FR. Conceicao Rodrigues College Of Engineering

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

Department of Computer Engineering

B.E. (Computer) (semester VIII) (2019-2020)
Course Outcomes & Assessment Plan

Subject: Natural Language Processing (NLP-DLO8012) **Credits-4+1**

PAC Members:	**H.O.D.**

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
4. To implement various language Models.
5. To design systems that uses NLP techniques

Syllabus:

Prerequisite: Data structure & Algorithms, Theory of computer science, Probability Theory.

Module No.	Unit No.	Topics	Hrs.
1	**Introduction**	History of NLP, Generic NLP system, levels of NLP , Knowledge in language processing , Ambiguity in Natural language , stages in NLP, challenges of NLP ,Applications of NLP	4
2	**Word Level Analysis**	Morphology analysis –survey of English Morphology, Inflectional morphology & Derivational morphology, Lemmatization, Regular expression, finite automata, finite state transducers (FST) ,Morphological parsing with FST , Lexicon free FST Porter stemmer. N –Grams- N-gram language model, N-gram for spelling correction.	10
3	**Syntax analysis**	Part-Of-Speech tagging(POS)- Tag set for English (Penn Treebank) , Rule based POS tagging, Stochastic POS tagging, Issues –Multiple tags & words, Unknown words. Introduction to CFG, Sequence labeling: Hidden Markov Model (HMM), Maximum Entropy, and Conditional Random Field (CRF).	10
4	**Semantic Analysis**	Lexical Semantics, Attachment for fragment of English- sentences, noun phrases, Verb phrases, prepositional phrases, Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy, WordNet, Robust Word Sense Disambiguation (WSD) ,Dictionary based approach	10

5	**Pragmatics**	Discourse –reference resolution, reference phenomenon , syntactic & semantic constraints on co reference	8
6	**Applications (preferably for Indian regional languages)**	Machine translation, Information retrieval, Question answers system, categorization, summarization, sentiment analysis, Named Entity Recognition.	10

Text Books:

1. Daniel Jurafsky, James H. Martin "Speech and Language Processing" Second Edition, Prentice Hall, 2008.
2. Christopher D.Manning and Hinrich Schutze, " Foundations of Statistical Natural Language Processing ", MIT Press, 1999.

Reference Books:

1. Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press (2008).
2. Daniel M Bikel and Imed Zitouni " Multilingual natural language processing applications" Pearson, 2013
3. Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor) " The Handbook of Computational Linguistics and Natural Language Processing " ISBN: 978-1-118-
4. Steven Bird, Ewan Klein, Natural Language Processing with Python, O'Reilly
5. Brian Neil Levine, An Introduction to R Programming
6. Niel J le Roux, Sugnet Lubbe, A step by step tutorial : An introduction into R application and programming

Assessment:

Internal Assessment:
Assessment consists of two class tests of 20 marks each. The first class test is to be conducted when approx. 40% syllabus is completed and second class test when additional 40% syllabus is completed. Duration of each test shall be one hour.

End Semester Theory Examination:
1 Question paper will comprise of 6 questions, each carrying 20 marks.
2. The students need to solve total 4 questions.
3. Question No.1 will be compulsory and based on entire syllabus.
4. Remaining question (Q.2 to Q.6) will be selected from all the modules.

Teaching Scheme

Course Code	Course Name	Teaching Scheme			Credits Assigned			
		Theory	Practical	Tutorial	Theory	Practical/ Oral	Tut	Credits
DLO8012	Natural Language Processing	04	--	--	04	--	---	04
CSL804	Computational Lab-II	--	02	--	--	1	--	01

Examination Scheme

Course Code	Course Name	Examination Scheme							
		Theory Marks				Term Work	Practical	Oral	Total
		Internal Assessment			End Sem Exam				
		Test1	Test2	Avg					
DLO8012	Natural Language Processing	20	20	20	80 (3hr)	--	---	--	100
CSL804	Computational Lab-II					50	--	25	75

FR. Conceicao Rodrigues College Of Engineering

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

Department of Computer Engineering

B.E. (Computer) (semester VIII) (2019-2020)

Lecture Plan:

Subject: Natural Language Processing (NLP-DLO8012) **Credits-04**

Time Table (2 week):

	Prof. Swati Ringe						**With Effect from 01st Jan 2019**				
	8.45-9.30	**9.30-10.15**	**10.15-11.00**	**11.00-11.15**	**11.15-12.00**	**12.00-12.45**	**12.45-13.15**				
Mon				**BREAK**			**LUNCH**				
Tues			**NLP BEC**								
Wed	**NLP BEC**										
Thurs			**NLP BEC**								
Fri	**NLP BEC**										

Time Table (Regular):

Prof. Swati Ringe							**With Effect from 20th Jan 2020**			
	8.45-9.45	**9.45-10.45**	**10.45-11.00**	**11.00-12.00**	**12.00-01.00**	**13.00-13.30**	**13.30-14.30**	**14.30-15.30**	**15.30-16.30**	**16.30-17.30**
Mon			**BREAK**	**OSL (SEC-C)**		**LUNCH**				
Tues							**NLP BEC**	**NLP (BEC-A)**		
Wed	**NLP BEC**			**NLP (BEC-D)**			**NLP BEC**	**OSL (SEC-C)**		
Thurs				**NLP (BEC-C)**						
Fri				**NLP (BEC-B)**			**NLP BEC**			

Total Load: 4T + 12P = 16 + MENTOR

Lecture Plan : SEM VIII-NLP-DLO8012

Modes of Content Delivery:

i	Class Room Teaching	v	Self Learning Online Resources	ix	Industry Visit
ii	Tutorial	vi	Slides	x	Group Discussion
iii	Remedial Coaching	vii	Simulations/Demonstrations	xi	Seminar
iv	Lab Experiment	viii	Expert Lecture	xii	Case Study

No.	Portion to be covered	Planned date	Actual date	Content Delivery - Reference /Assessment Method
1.	**Introduction :** History of NLP.	**07/01/2020**		Slides/TB1-2/NPTEL/UT1
2	Generic NLP system	**08/01/2020**		Slides/TB1-2/NPTEL/UT1
3	levels of NLP.	**09/01/2020**		Slides/TB1-2/NPTEL/UT1
4	Knowledge in language processing	**10/01/2020**		Slides/TB1-2/NPTEL/UT1
5	Ambiguity in Natural language	**14/01/2020**		Slides/TB1-2/NPTEL/UT1
6	stages in NLP	**15/01/2020**		Slides/TB1-2/NPTEL/UT1
7	challenges of NLP ,	**16/01/2020**		Slides/TB1-2/NPTEL/UT1
8	Applications of NLP	**17/01/2020**		Slides/TB1-2/NPTEL/UT1
9	**Word Level Analysis:** Morphology analysis , survey of English Morphology	**21/01/2020**		Slides/TB1-2/NPTEL/UT1
10	Inflectional morphology	**22/01/2020**		Slides/TB1-2/NPTEL/UT1
11	Derivational morphology	**23/01/2020**		Slides/TB1-2/NPTEL/UT1
12	Lemmatization	**24/01/2020**		Slides/TB1-2/NPTEL/UT1
13	Regular Expression and finite automata	**28/01/2020**		Slides/TB1-2/NPTEL/UT1
14	finite state transducers (FST) Morphological parsing with FST	**29/01/2020**		Slides/TB1-2/NPTEL/UT1
15	Lexicon free FST Porter stemmer.	**30/01/2020**		Slides/TB1-2/NPTEL/UT1

16	N Grams- N-gram language model, N-gram for spelling correction	**31/01/2019**		Slides/TB1-2/NPTEL/UT1
17	**Syntax analysis:** Part-Of-Speech tagging(POS)- Tag set for English (Penn Treebank)	**04/02/2020**		Slides/TB1-2/NPTEL/UT1
18	Rule based POS tagging, Stochastic POS tagging,	**05/02/2020**		Slides/TB1-2/NPTEL/UT1
19	Issues Multiple tags & words, Unknown words.	**06/02/2020**		Slides/TB1-2/NPTEL/UT1
20	Introduction to CFG.	**07/02/2020**		Slides/TB1-2/NPTEL/UT1
21	Sequence labeling: Hidden Markov Model (HMM),	**11/02/2020**		Slides/TB1-2/NPTEL/UT2
22	Maximum Entropy	**12/02/2020**		Slides/TB1-2/NPTEL/UT2
23	Conditional Random Field (CRF).	**13/02/2020**		Slides/TB1-2/NPTEL/UT2
24	Semantic Analysis : Lexical Semantics	**14/02/2020**		Slides/TB1-2/NPTEL/UT2
25	Attachment for fragment of English-sentences,	**18/02/2020**		Slides/TB1-2/NPTEL/UT2
26	noun phrases, Verb phrases, prepositional phrases	**25/02/2020**		Slides/TB1-2/NPTEL/UT2
27	Relations among lexemes & their senses Homonymy	**03/03/2020**		Slides/TB1-2/NPTEL/UT2
28	Polysemy	**04/03/2020**		Slides/TB1-2/NPTEL/UT2
29	Synonymy	**05/03/2020**		Slides/TB1-2/NPTEL/UT2
30	Hyponymy	**06/03/2020**		Slides/TB1-2/NPTEL/UT2
31	WordNet	**11/03/2020**		Slides/TB1-2/NPTEL/UT2
32	Robust Word Sense Disambiguation (WSD)	**12/03/2020**		Slides/TB1-2/NPTEL/UT2
33	Dictionary based approach	**13/03/2020**		Slides/TB1-2/NPTEL/UT2
34	Pragmatics: Discourse reference resolution	**17/03/2020**		Slides/TB1-2/NPTEL/UT2
35	reference phenomenon	**18/03/2020**		Slides/TB1-2/NPTEL/UT2
36	syntactic constraints on co reference	**19/03/2020**		Slides/TB1-2/NPTEL/UT2
37	semantic constraints on co reference	**20/03/2020**		Slides/TB1-2/NPTEL/UT2
38	Applications (preferably for Indian regional languages): Machine translation	**24/03/2020**		Slides/TB1-2/NPTEL/UT2
39	Information retrieval	**26/03/2020**		Slides/TB1-2/NPTEL/UT2

40	Question answers system	**27/03/2020**		Slides/TB1-2/NPTEL/UT2
41	Categorization	**31/03/2020**		Slides/TB1-2/NPTEL/UT2
42	Summarization	**01/04/2020**		Slides/TB1-2/NPTEL/UT2
43	sentiment analysis	**02/04/2020**		Slides/TB1-2/NPTEL/UT2
44	Named Entity Recognition.	**03/04/2020**		Slides/TB1-2/NPTEL/UT2

Total Lectures : 44

Reference

Text Book 1 : TB1
Natural Language Processing – STAREDU Solutions : TB2
Slides

Reference Web Resources:

1. NPTEL Course on Natural Language Processing
2. Edureka NLP Course

Course Outcomes:

On successful completion of course learner should be able to

DLO8012.1: Have a broad understanding of the field of natural language processing.

DLO8012.2: Have a sense of the capabilities and limitations of current natural language technologies.

DLO8012.3: Be able to model linguistic phenomena with formal grammars.

DLO8012.4: Be able to Design, implement and test algorithms for NLP problems.

DLO8012.5: Understand the mathematical and linguistic foundations underlying approaches to the various areas in NLP.

DLO8012.6: Be able to apply NLP techniques to design real world NLP applications such as machine translation, text categorization, text summarization, information extraction.

Mapping of CO and PO/PSO

Relationship of course outcomes with program outcomes: Indicate 1 (low importance), 2 (Moderate Importance) or 3 (High Importance) in respective mapping cell.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
DLO8012.1	3												3	
DLO8012.2	3	3											3	3
DLO8012.3	3	3	3		2				2				3	3
DLO8012.4	3	3	3	3	3				2				3	3
DLO8012.5	3	3	3	2	3				2			2	3	3
DLO8012.6	3	3	3	2	3				3	3	2	3	3	3
TOTAL	18	15	12	7	11				9	3	2	5	18	15
CO-PO MATRIX	**3**	**3**	**3**	**2.3**	**2.75**				**2.25**	**3**	**2**	**2.5**	**3**	**3**

Course Outcomes Target: [Target 2.5]

Course Outcomes:

On successful completion of course learner should be able to

DLO8012.1: Have a broad understanding of the field of natural language processing.

DLO8012.2: Have a sense of the capabilities and limitations of current natural language technologies.

DLO8012.3: Be able to model linguistic phenomena with formal grammars.

DLO8012.4: Be able to Design, implement and test algorithms for NLP problems.

DLO8012.5: Understand the mathematical and linguistic foundations underlying approaches to the various areas in NLP.

DLO8012.6: Be able to apply NLP techniques to design real world NLP applications such as machine translation, text categorization, text summarization, information extraction.

CO ASSESSMENT TOOLS

	Direct Methods (80%)						**Indirect Methods (20%)**
DLO8012.1	Test1 (60%)		UE –Th (20%)	UE-O (20%)			(100%)
DLO8012.2	Test1 (30%)	Lab 1-2-3-4-5-8 (40%)	UE -Th (20%)	UE-O (10%)			(100%)
DLO8012.3	Test1 (30%)	Lab 6 (10%)	UE –Th (20%)	UE-O (20%)	Assign2 (20%)		(100%)
DLO8012.4	Test2 (30%)	Lab 7 (10%)	UE –Th (20%)	UE-O (20%)	MP (20%)		(100%)
DLO8012.5	Test2 (30%)	Assign3 (10%)	UE –Th (20%)	UE-O (20%)	MP (20%)		(100%)
DLO8012.6	MP (50%)	Assign 1 (30%)	UE –Th (10%)	UE-O (10%)			(100%)

Content Beyond Syllabus:

1. Research Paper study individually.

Curriculum Gap:

The students need to know basics of Probability Theory.

In order to achieve the course objectives, there are some topics listed below are not given much importance.

Sr.No.	Content Beyond Syllabus	Action Plan	PO Mapping
1	Probability Theory	Planned one lecture.	PO2, PSO2

Department of Computer Engineering
Academic Term: Jan-April 2020

Rubrics for Lab Experiments

Class : B.*E. Computer* **Subject Name :*NLP***
Semester : VIII **Subject Code :*DLO8012***

Practical No:	
Title:	
Date of Performance:	
Roll No:	
Name of the Student:	

Evaluation:

Indicator	**Very Poor**	**Poor**	**Average**	**Good**	**Excellent**
Timeline (2)	More than three sessions late (0)	More than two sessions late (0.5)	Two sessions late (1)	One session late (1.5)	Early or on time (2)
Efforts(3)	N/A	N/A	Not Completed (1)	Partially Completed (2)	Completed(3)
Legibility(3)	N/A	N/A	Poor(1)	Good(2)	Very Good(3)
PostLab(2)	N/A	N/A	N/A	Partially Correct(1)	All Correct(2)

Total Marks :
Signature of the Teacher :

Department of Computer Engineering
Academic Term : Jan-April 2020

Rubrics for Assignments

Class : B.*E. Computer* **Subject Name :*NLP***
Semester : VIII **Subject Code :*DLO8012***

Assignment No:	
Title:	
Date of Performance:	
Roll No:	
Name of the Student:	

Rubrics for Assignment Grading:

Indicator	**Very Poor**	**Poor**	**Average**	**Good**	**Excellent**
Timeline (2)	More than three sessions late (0)	More than two sessions late (0.5)	Two sessions late (1)	One session late (1.5)	Early or on time (2)
Organization (3)	N/A	Very poor readability and not structured (0.5)	Poor readability and somewhat structured (1)	Readable with one or two mistakes and structured (2)	Very well written and structured without any mistakes (3)
Level of content (3)	N/A	Major points are omitted or addressed minimally (0.5)	All major topics are covered, the information is accurate.(1)	Most major and some minor criteria are included. Information is Accurate (2)	All major and minor criteria are covered and are accurate. (3)
Depth of Knowledge(2)	N/A	One answer correct(0.5)	Two answers correct(1)	Three answers correct(1.5)	Four answers correct(2)

Total Marks :
Signature of the Teacher :

Department of Computer Engineering
Academic Term: Jan-April 2020
Rubrics for Mini Project

Class : B.*E. Computer* **Subject Name :*NLP***
Semester : VIII **Subject Code :*DLO8012***

Practical No:	
Title:	
Date of Performance:	
Roll No:	
Name of the Student:	

Rubric for Mini Project

Indicator	**Very Poor**	**Poor**	**Average**	**Good**	**Excellent**
Timeline: **Maintains project deadline (2)**	Project not done (0)	More than two session late (0.5)	Two sessions late (1)	One session late (1.5)	Early or on time (2)
Completeness: **Complete all parts of project (2)**	N/A	< 40% complete (0.5)	~ 60% complete (1)	~ 80% complete(1.5)	100% complete(2)
Application design: **(4)**	Design aspects are not used (0)	Poorly designed (1)	Project with limited functionalities (2)	Working project with good design (3)	Working project with good design and advanced techniques are used (4)
Presentation(2)	Not submitted report (0)	Poorly written and poorly kept report(0.5)	Report with major mistakes(1)	Report with less than 3-4 mistakes (1.5)	Well written accurate report(2)

Total marks: **Signature of Teacher:**

List of Experiments/Mini Project Plan

Expt No.	Batch A Tue	Batch B Fri	Batch C Thu	Batch D Wed	CO Map	Title/aim
01	21 Jan	24 Jan	23 Jan	22 Jan	CO2	Word Analysis.
02	28 Jan	31 Jan	30 Jan	29 Jan	CO2	Word Generation.
03	04 Feb	07 Feb	06 Feb	05 Feb	CO2	Stop Word Removal
04	11 Feb	14 Feb	13 Feb	12 Feb	CO2	Stemming
05	03 Mar	06 Mar	05 Mar	04 Mar	CO2	Morphology
06	17 Mar	13 Mar	12 Mar	11 Mar	CO3	POS tagging
07	24 Mar	20 Mar	19 Mar	18 Mar	CO4	Chunking
08	31 Mar	27 Mar	26 Mar	01 April	CO2	N-gram Language Model
09	**Mini Project:**				**CO6**	Implement one real life NLP application.(Group of 2/3).
	28 Jan					**Topic Submission**
	17 Mar					**Progress review**
	24 Mar					**Presentation and Demo**
	13 April					**Mini Project Report submission**

Assignments Plan

01	09 March	CO6	Industry or Domain specific Case Study of one Natural Language Processing Application
02	16 March	CO3	Discuss reference resolution problem with suitable example. Explain Lappin and Leass's algorithm for pronoun resolution with suitable example.
03	26 March	CO5/ PO12	Research Papers study.