

Branch Name:	Artificial Intelligence and Data Science
Program Code:	ET108
Semester/Year:	Semester VII/ Fourth Year
Course Name:	Natural Language Processing
Course Code:	1ET1080703
Pre-requisite Course:	Probability and statistics, Programming and data structures.

Course Objectives:

1. The Natural Language Processing course gives a detailed look at the science of applying machine learning algorithms to process large amounts of natural language data.
2. NLP is driving the growth of the AI market, and this course helps to develop the skills required to become an NLP Engineer.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)				
					Theory (Marks)		Practical (Marks)		Total (Marks)
	L	T	P	Total	University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
5	4	-	2	5	70	30	30	20	150

Subject Contents

Sr. No	Topic	Total Hours	Weight (%)
1	Introduction to NLP: What is NLP? Why NLP is Difficult? History of NLP, Advantages of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, How to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries	6	14
2	Language Modeling and Part of Speech Tagging: Unigram Language Model, Bigram, Trigram, N-gram, Advanced smoothing for language modeling, Empirical Comparison of Smoothing Techniques, Applications of Language Modeling, Natural Language Generation, Parts of Speech Tagging, Morphology, Named Entity Recognition	12	28
3	Words and Word Forms: Bag of words, skip-gram, Continuous Bag-Of-Words, Embedding representations for words Lexical Semantics, Word Sense Disambiguation, Knowledge Based and Supervised Word Sense Disambiguation	6	14
4	Text Analysis, Summarization and Extraction: Sentiment Mining, Text Classification, Text Summarization, Information Extraction, Named Entity Recognition, Relation Extraction, Question Answering in Multilingual Setting; NLP in Information Retrieval, Cross-Lingual IR	10	22

5	Machine Translation: Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM), Encoder-decoder architecture, Neural Machine Translation	10	22
---	--	----	----

List of Reference Books:

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jurafsky, David, and James H. Martin, PEARSON
2. Foundations of Statistical Natural Language Processing, Manning, Christopher D., and Hinrich Schütze, Cambridge, MA: MIT Press.
3. Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc..
4. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit Steven Bird, Ewan Klein, and Edward Loper.

List of Experiments:

1. Install NLTK
2. Implementation of Text Pre-Processing using NLTK (Tokenization, Stemming and Lemmatization and removal of stop words) in NLP
3. Write a program to convert text to speech
4. Write a program to audio file to text
5. Study of various Corpus-Brown, Inaugural udhr with various methods like fields, raw, words, sents
6. Study about conditional frequency distributions
7. Write a program to find the most frequent noun tags
8. Map words to properties using python dictionaries
9. Study about DefaultTagger, Regular expression tagger, UnigramTagger
10. Write a program to create and use your own corpora (plaintext, catogarical)

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand comprehend the key concepts of NLP and identify the NLP challenges and issues.	1. Remembering 2. Understanding
CLO2	Develop Language Modeling for various text corpora across the different languages	4. Analyze 5. Evaluate
CLO3	Illustrate computational methods to understand language phenomena of word sense disambiguation	4. Analyze
CLO4	Design and develop applications for text or information extraction/summarization/classification	2. Understanding 4. Analyze
CLO5	Apply different Machine translation techniques for translating a source to target language(s)	3. Applying 6. Creating

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1			M	L	M	L						M	H	M	L	
CLO2	M	H	M	M	H	L				L		M	M	M	L	
CLO3	M	M	H	M	M	L							H	H	M	L
CLO4	L	M	M	M	M			L					M		M	H
CLO5	H	H	H	M	H				L			H	M		M	H

H: High, M: Medium, L: Low