National Institute of Technology Meghalaya
An Institute of National Importance

CURRICULUM

Programme	**Master of Computer Applications**		Academic Year of Regulation	**2024-25**
Department	**Computer Science and Engineering**		Semester	**IV**

Course Code	Course Name	Pre-Requisite	Credit Structure				Marks Distribution			
			L	T	P	C	INT	MID	END	Total
CA576	**Natural Language Processing**		**3**	**0**	**0**	**3**	**50**	**50**	**100**	**200**

Course Objectives		Course Outcomes	CO's	Statement	Bloom's Taxonomy
	This course introduces foundational linguistic and mathematical concepts and algorithms for analysis of natural languages.		CA576.1	Able to understand the techniques for basic linguistic processing for phonetic analysis, phonological analysis and morphological analysis.	Understand
	This course introduces the advantages and disadvantages of different NLP technologies in different real-life applications.		CA576.2	Able to design computational models of natural language text data in order to gain broader understanding of text data.	Create
	This course familiarizes some statistical approaches and machine learning techniques used in Natural Language Processing (NLP) tasks.		CA576.3	Able to evaluate common NLP tasks using models, methods, and algorithms for statistical NLP.	Evaluate
			CA576.4	Able to create software implementations of relevant pre-processing steps for different NLP problems.	Create
			CA576.5	Able to Evaluate and implement common NLP tasks using machine learning algorithms.	Evaluate

COs	Mapping with Program Outcomes (POs)												Mapping with PSOs		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CA576.1	3	1	1				1					1			
CA576.2	3	3	3	2	2								1	1	1
CA576.3	3	3	3	2	2	1			1		1			1	1
CA576.4	3	3	3	2	1	1								2	3
CA576.5	3	3	3	2	2	1			1	1	1			1	1
CA576	3	2.6	2.6	1.6	1.4	0.6	0.2		0.4	0.2	0.4	0.2	0.2	1	1.2

SYLLABUS

No.	Content	Hours	COs
I	Introduction; Motivation and challenges of Natural Language Processing (NLP); Tokenisation and Sentence Segmentation	04	CA576.1
II	Lexical Analysis: Morphology, Finite State Morphology	04	CA576.1, CA576.2
III	Syntactic Analysis: Linguistic Background - An outline of English Syntax, Grammars for Natural Language, Parsing techniques, Linking Syntax and Semantics; Semantic Analysis: Lexical Semantics, Word Sense Disambiguation; Pragmatics and Discourse Analysis: Dialogue and Conversational agents, Co-reference resolution; Natural Language Generation	12	CA576.2, CA576.3
IV	Overview of NLP applications: POS tagging, Information Retrieval, Question Answering, Information Extraction, Dialogue Systems, Text and Intent Mining, Machine Translation; Data pre-processing for NLP tasks	12	CA576.3
V	Empirical techniques for NLP tasks; machine learning techniques for NLP tasks; NLP application examples in real-life; Performance evaluation metrics for NLP systems	10	CA576.4, CA576.5
	Total Hours	42	

Essential Readings

1. D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition," Pearson Education India, 2nd edition, 2013.

2. Akshar Bharati, Vineet Chaitanya, Rajeev Sangal, "Natural Language Processing: A Paninian Perspective", PHI Learning Pvt. Ltd., 1st edition, 1995.

3. Daniel M. Bikel, "Multilingual Natural Language Processing Applications: From Theory to Practice", Pearson Education India, 1st edition, 2012.

4. C. D. Manning, H. Schütze, "Foundations of Statistical Natural Language Processing", MIT Press, 1st edition, 1999.

Supplementary Readings

1. Jacob Perkins, "Python 3 Text Processing with NLTK 3 Cookbook", Packt Publishing Limited, 1st edition, 2014.

2. Breck Baldwin, Krishna Dayanidhi, "Natural Language Processing with Java and LingPipe Cookbook", Packt Publishing Limited, 1st edition, 2014.

3. Nitin Indurkhya and Fred J. Damerau, "Handbook of Natural Language Processing", Taylor and Francis, 2nd edition, 2010.