

Name: Manish Shrivastava
Course: CSE
Email: m.shrivastava@iiit.ac.in
Mobile: 7032900342

Name of the Academic Program: **B.Tech. in Computer Science and Engineering**

Course Code: **CS7.401** Title of the Course: **Introduction to NLP**

L-T-P: **3-1-0**. Credits: **4**

(L= Lecture hours, T=Tutorial hours, P=Practical hours)

1. Prerequisite Course / Knowledge:

None

2. Course Outcomes (COs)

After completion of this course successfully, the students will be able to –

- CO-1. Demonstrate the knowledge of stages and fundamental building blocks of NLP
- CO-2. Apply NLP machine learning algorithms for classification, representation, and parsing
- CO-3. Demonstrate the knowledge of Dense vector representation for NLP
- CO-4. Explain the concepts behind distributed semantics
- CO-5. Discuss the approaches to global and contextual semantic representation
- CO-6. Apply the above concepts for fundamental NLP tasks.

3. Mapping of Course Outcomes (COs) with Program Outcomes (POs) and Program Specific Outcomes (PSOs) – Course Articulation Matrix

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	1	3	3	2	2	2	1	-	1	1	1	3	3	3	3	2
CO2	3	3	3	3	3	1	1	1	3	3	3	3	3	3	2	3
CO3	3	3	3	3	3	-	-	-	2	2	2	3	3	3	2	3
CO4	3	3	3	3	3	-	-	-	2	2	2	3	3	3	2	3
CO5	3	3	3	3	3	-	-	-	2	2	2	3	3	3	2	3
CO6	3	3	3	3	3	1	1	1	3	3	3	3	3	3	3	3

‘3’ in the box denotes ‘High-level’ mapping, 2 for ‘Medium-level’ mapping, 1 for ‘Low’-level’ mapping

4. Detailed Syllabus:

Unit 1: Stages of NLP: from lexical to semantic. Fundamental Language processing : Tokenization, Language modeling, Text classification,

Unit 2: Morphology, POS Tagging, Chunking, Discriminative vs generative modes, HMM and CRF

Unit 3: Syntax parsing: Constituency and Dependency, PCFG, projectivity Arc-eager

Unit 4: Distributed semantics: SVD, Word2Vec, RNN, LSTM,

Unit 5: Contextual Distributed semantics: ELMO, BERT

- Four assignments related to the above syllabus will be done by students.
- One course project

References:

Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.

5. Teaching-Learning Strategies in brief:

Lectures by integrating ICT into classroom teaching, weekly tutorials involving problem solving and active learning by students and Project-based Learning by doing four assignments.

6. Assessment methods and weightages in brief :

- Assignments: 20%
- Mid Evaluation (written): 10 %
- Project: 40%
- End Evaluation (Viva): 20%
- Quiz: 10%