KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
Deemed to be University
BHUBANESWAR-751024

School of Computer Engineering
Spring Semester 2025

Natural Language Processing (NLP)-CS30016 (L-T-P-Cr:3-0-0-3)

Instructor Name: Dr. Sambit Praharaj
Email: sambit.praharajfcs@kiit.ac.in
Cabin: B 221(D), Campus 25

Course Objectives:

- To understand the steps involved in Natural language processing
- To learn about the lexical, syntactic and semantic analysis of natural language processing
- To explore the various parsing techniques for natural languages
- To understand the statistical models for Natural language processing
- To learn about the various applications involved in Natural language processing.

Day-Wise Lesson plan:-

Lecture no.	**Topics to be covered**
1	• Introduction to NLP (With real-world examples)
2	• Phases of NLP (Lexical and Morphological, Syntactic, Semantic, Discourse Integration, Pragmatic) • NLP Pipeline
3	• NLP Applications • Latest NLP Tools • Demo with NLTK

4	● Why is NLP hard? Ambiguity in NLP, Various types of ambiguity with examples ● Empirical Laws (Zip's Law and Heap's Law) (Numerical and Conceptual Examples)
5	● What is text processing? Need for normalization and cleaning: Unicode, encoding, whitespace, punctuation, case-folding ● Regular Expression and String Matching, Pattern matching
6	● Phonology and Morphology ● Phonemes, syllables (phonology basics) ● Morphs, morphemes, affixes ● Word formation ● Why does morphological understanding matter for stemming/ lemmatization?
7	● Tokenization ● Sentence Segmentation ● Word tokenization ● Subword tokenization (Intro to BPE and WordPiece with simple numerical examples) ● Stemming and Lemmatization ● Stemming algorithms (Intro to Porter Stemmer) ● Lemmatization ● Stop Words Removal
8	● Part-of-Speech Tagging (POS: NN, VB, JJ, RB, etc.) ● Intro to Rule-based, HMM-based, Neural POS taggers ● Named Entity Recognition (NER) ● Intro to Sequence labeling
9	● Language Modeling: Bi-gram, Tri-gram, N-gram Model (numerical examples of probabilistic models) ● Intro to Noisy Channel Model and Smoothing (only basic Laplace and primary Backoff smoothing with numerical examples)
10	● Basics of Spelling Correction. (Numerical examples with Edit Distance-Levenshtein distance only)
11	● What is the word Feature Extraction and Representation? Word Embedding (Text to Vector) ● BoW, CBoW
12	● TF-IDF (Conceptual and numerical examples)

	• Intro to Word2Vec with examples (CBoW and Skipgram intro with examples)
13	• Hands-on Exercise on Text Pre-Processing (Using NLTK, Spacey)
14	• Text Classification: Naive Bayes and Sentiment Classification • Naive Bayes Classifier
15	• Training the Naive Bayes Classifier
16	• Hands-on Exercise on Naive Bayes Classifier
17	• Other Text Classification tasks using Naive Bayes
18	• Evaluation: Confusion Matrix, Precision, Recall, F-Measure
19	• Hands-on Exercise on Evaluation of Text Classifier
20	• Tutorials/ Activity/ Project/ Assignment: NLP Applications: Chatbot, Classifier, Spam Detector, Sentiment Analysis from X or Social Media, Projects based on NLP Pipeline
MID SEMESTER	
21	• Markov Models, Hidden Markov Models (HMMs)
22	• Viterbi Algorithm
23	• Estimating the parameters of HMMs
24	• The Forward-Backward Algorithm
25	• Implementation Issues
26	• HMM Tagging, Word Segmentation and NER
27	• Intro to RNN (Learning meaning of words, not just probabilities) • Advantages of RNN over HMM. Comparison

28	• RNN: Recurrent Hidden State (Dependency Learning)
29	• RNN: Gating Mechanisms (LSTM/ GRU) for long term dependency learning
30	• Attention Mechanism
31	• Self Attention
32	• Multi-head attention
33	• Transformers Encoder (BERT)
34	• Transformers Decoder (Cross Attention)
35	• Encoder-Decoder Architecture
36	• Intro to Speech and its role in NLP (Biology of Human Speech Production, From Waveform to Spectrogram) • Speech vs Text in NLP
37	• Speech Signal Processing Essentials (Framing and Windowing, Frequency Domain Basics, Speech Features)
38	• Feature Engineering vs Deep Learning (MFCC vs wav2vec features)
39	• Intro to Automatic Speech Recognition Pipeline with examples
40	• Hands-on Exercise on Speech Processing
Tutorials / Activity/ Project/ Assignment: Transformer based models, RNN, LSTM, GRU, BERT, Speech Processing.	

Course Outcome:

At the end of the course, the students will be able to:

CO1: Understand the concepts of NLP and its algorithms.
CO2: Evaluate different computing architectures for natural language processing for various parameters.

CO3: Apply the different modelling and tagging concepts for language processing.
CO4: Analyze the various grammars & parsing algorithms.
CO5: Explore the role of statistical parsing & machine translation.
CO6: Implementation of the role of natural language processing in real life applications.

Activity Calendar:

Name	**Tentative Date**	**Marks**
Conceptual Assignment	17 Jan 2026	5
Class Test (Syllabus till Lecture 12)	19/ 20 Jan 2026	5
Programming Assignment 1	25 Jan 2026	5
Programming Assignment 2	13 March 2026	5
Class Notes Check	23/ 24 Mar 2026	5
Class Performance and Attendance	27 Mar 2026	5

Text Books:

1. Christopher Manning, Schutze Heinrich, “Foundations of Statistical Natural Language Processing”, MIT Press, 1999.
2. Speech and Language Processing, Jurafsky, D. and J. H. Martin, Prentice-Hall.

Reference Books:

1. Natural Language Understanding, Allen, J., The Benjamins/Cummings Publishing Company Inc.
2. Statistical Methods for Speech Recognition, Jelinek, F., The MIT Press.

Online Resources:

1. NPTEL Course by Prof. Pawan Goyal: https://nptel.ac.in/courses/106105158
2. Community curated list of NLP Resources: https://github.com/keon/awesome-nlp
3. High Quality Resources for NLP: https://medium.com/nlplanet/awesome-nlp-18-high-quality-resources-for-studying-nlp-1b4f7fd87322
4. Stanford NLP Course: https://web.stanford.edu/class/cs224n/