

Natural Language Processing (CSET 346)

Course Type : Specialization Elective

Program : B.Tech

Semester and Year : V Semester & 3rd Year

L-T-P : 2 – 0 – 2

Credits : 3

School : SCSET

Level : UG (B.Tech)

School of Computer Science Engineering and Technology



Bennett University
Greater Noida, Uttar Pradesh

COURSE BRIEF

COURSE TITLE	Natural Language Processing	PRE-REQUISITES	NA
COURSE CODE	CSET346	TOTAL CREDITS	3
COURSE TYPE	Specialization Elective	L-T-P FORMAT	2-0-2

COURSE SUMMARY

This course aims to provide a broad survey of Natural Language Processing (NLP) in both Indian and international context. This course examines NLP models and algorithms using both traditional, symbolic recent machine learning and deep learning-based approaches. At the end of this course, the students will be able to implement state-of-the-art Machine Learning and Deep Learning solutions to NLP problems in Global & Indian context.

COURSE-SPECIFIC LEARNING OUTCOMES (CO)

By the end of this program, students should have the following knowledge, skills and values:

CO1: To articulate natural language processing and importance of word representation.

CO2: To build deep learning model for solving natural language problems such as language modelling, machine translation, text summarization, POS tagging, Seq2Seq generation.

CO3: To implement state-of-the-art Machine Learning and Deep Learning solutions to NLP problems in Global & Indian context.

Detailed Syllabus

Module 1 (Contact hours: 7)

Natural Language Processing: Need, applications, industry demand, Challenges in NLP: Ambiguity in language, Contextual words and phrases and homonyms, Coreference, Domain-specific language, Low-resource languages, Text Pre-processing techniques: Segmentation, Stemming, Lemmatization, Spelling correction, Synsets, Hypernyms, Tokenization, N-grams, Stops Words, WordNet, WordNet Similarity, Language Corpus, N-gram Language Models, Hidden Markov Models.

Module 2 (Contact hours: 7)

Topic Modelling, Latent Semantic Analysis, Text normalisation, Parallel Corpus, Handling Code-mix text, Word representation, Sentence representation, Word embedding, Term Frequency, TF-IDF Representation, Distributional representation, Word2vec: CBOW, Word embedding for regional language, Word2Vec, GloVe, Document to Vector.

Module 3 (Contact hours: 7)

Neural Networks for text, Recurrent Neural Networks, Vanishing Gradients, exploding gradient, LSTM (Long short term memory), GRU (Gated recurrent Unit), Seq2Seq Modelling, Bidirectional Model, Contextual Representations, Transformers, BERT, Transfer Learning, POS tagging, Named Entity Recognition, Sentiment Analysis.

Module 4 (Contact hours: 7)

Self-Attention mechanism, Cross attention, Question Answering Bot, 1D-CNN for NLP, Sub-word models, Open AI's GPT and its variants, ULMFiT, Facebook's RoBERTa, Meta's LLaMa model, Low-Rank Adaptation of Large Language Models (LoRA), Application area: Text Summarization, Extractive, Abstractive Text summarization, Transformer models for Text Summarization.

STUDIO WORK / LABORATORY EXPERIMENTS:

Implementation of text document reading, parsing and applying text pre-processing approaches. Implement wordnet based document representation and word searching. Implement word level and sentence level text representation for searching. Implement one hot vector of the given dataset (Indian language). Design first neural network for text processing.. Implement topic modelling using clustering and LSA. Implement text classification model for sentiment analysis. Implement Part of Speech (POS) tagging and Named Entity Recognition (NER). Implement Machine Translation model. Implement chatbot using sequence to sequence modelling. Implement Text Generation models for social media, news context.

TEXTBOOKS/LEARNING RESOURCES:

1. Delip Rao and Brian McMahan., Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning (1 ed.), O'ReillyMedia, 2019. ISBN 978-1491978238.

REFERENCE BOOKS/LEARNING RESOURCES:

1. Jacob Eisenstein, Introduction to Natural Language Processing (1 ed.), The MIT Press; Illustrated edition, 2019. ISBN 9780262042843.

2. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana, Practical Natural Language Processing (1 ed.), O'Reilly, 2020. ISBN 978149205402X

EVALUATION POLICY:

Components of Course Evaluation	Percentage
Mid Term Examination	20
End Term Examination	40
Certification	10
Project	10
Lab Continuous Evaluation	20

MOOC Mapping:

<https://www.coursera.org/specializations/natural-language-processing>

- Natural Language Processing with Classification and Vector Spaces
- Natural Language Processing with Probabilistic Models
- Natural Language Processing with Sequence Models
- Natural Language Processing with Attention Models.