

# MOOC APPROVAL REQUEST

As per KTU B.Tech Regulations 2024, Section 17

---

## KTU COURSE

Code: PECST862  
Name: Natural Language Processing

## NPTEL COURSE

Name: Natural Language Processing  
Instructor: Prof. Pawan Goyal  
Institution: IIT Kharagpur  
Duration: 12 Weeks  
Course ID: noc26-cs45

Semester: Jan-Apr 2026  
Date: December 02, 2025

### Document Contents:

1. KTU Course Syllabus (Complete)
2. NPTEL Course Details
3. Syllabus Comparison Report

## SEMESTER S8

### NATURAL LANGUAGE PROCESSING

(Common to CS/CA/CD)

<b>Course Code</b>	<b>PECST862</b>	<b>CIE Marks</b>	40
<b>Teaching Hours/Week (L:T:P: R)</b>	3:0:0:0	<b>ESE Marks</b>	60
<b>Credits</b>	3	<b>Exam Hours</b>	2 Hrs. 30 Min.
<b>Prerequisites (if any)</b>	None	<b>Course Type</b>	Theory

#### Course Objectives:

1. To provide a comprehensive understanding of natural language processing (NLP) and language models, focusing on the principles and techniques of prompt engineering to effectively guide and optimize AI-driven outputs.
2. practical skills necessary to design, implement, and evaluate prompt engineering strategies across various applications, while considering the ethical implications and challenges associated with AI-generated content.

#### SYLLABUS

<b>Module No.</b>	<b>Syllabus Description</b>	<b>Contact Hours</b>
<b>1</b>	<b>Introduction to NLP:</b> Introduction to Natural Language Processing - Various stages of traditional NLP – Challenges - Basic Text Processing techniques - Common NLP Tasks. N-gram Language Models - Naive Bayes for Text Classification, and Sentiment Analysis – Evaluation-Precision, Recall and F-measure-Test sets and cross validation.	<b>7</b>
<b>2</b>	<b>Traditional NLP Techniques:</b> Annotating Linguistic Structures - Context-Free Grammars, Constituency Parsing, Ambiguity, CYK Parsing, Dependency Parsing - Transition-Based Dependency Parsing, Graph-Based Dependency Parsing, Evaluation.	<b>7</b>
<b>3</b>	<b>Neural Networks for NLP:</b> Word representations - Lexical Semantics, Vector Semantics, TF-IDF, Pointwise Mutual Information (PMI), Neural Word embeddings - Word2vec, GloVe, Contextual Word Embeddings. Evaluating Vector Models - Feedforward Neural Networks for Text Classification	<b>10</b>

<b>4</b>	<b>Advanced NLP and Applications:</b> Sequence Modelling - Recurrent Neural Networks, RNNs as Language Models, RNNs for NLP tasks, Stacked and Bidirectional RNN architectures, Recursive Neural Networks, LSTM & GRU, Common RNN NLP Architectures, Encoder-Decoder Model with RNNs, Attention models, Transformers. NLP Applications - Machine Translation, Question Answering and Information Retrieval, Introduction to Large Language Models.	<b>12</b>
----------	--	-----------

**Course Assessment Method**  
**(CIE: 40 marks, ESE: 60 marks)**

**Continuous Internal Evaluation Marks (CIE):**

<b>Attendance</b>	<b>Assignment/ Microproject</b>	<b>Internal Examination-1 (Written)</b>	<b>Internal Examination- 2 (Written)</b>	<b>Total</b>
<b>5</b>	<b>15</b>	<b>10</b>	<b>10</b>	<b>40</b>

**End Semester Examination Marks (ESE)**

*In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions*

<b>Part A</b>	<b>Part B</b>	<b>Total</b>
<ul style="list-style-type: none"> <li>• 2 Questions from each module.</li> <li>• Total of 8 Questions, each carrying 3 marks</li> </ul> <p><b>(8x3 =24 marks)</b></p>	<ul style="list-style-type: none"> <li>• Each question carries 9 marks.</li> <li>• Two questions will be given from each module, out of which 1 question should be answered.</li> <li>• Each question can have a maximum of 3 subdivisions.</li> </ul> <p><b>(4x9 = 36 marks)</b></p>	<b>60</b>

## Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
<b>CO1</b>	Understand the foundational concepts of NLP and apply that to do text processing.	<b>K3</b>
<b>CO2</b>	Utilize word representations and evaluate vector models for NLP	<b>K3</b>
<b>CO3</b>	Analyse and implement advanced linguistic annotation and parsing techniques	<b>K4</b>
<b>CO4</b>	Apply advanced sequence modeling techniques using Neural Networks	<b>K3</b>
<b>CO5</b>	Apply NLP techniques in machine translation, question answering, and information retrieval.	<b>K3</b>

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

## CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
<b>CO1</b>	3	3								3		
<b>CO2</b>	3	3			3							
<b>CO3</b>	3	3									3	
<b>CO4</b>	3	3	3		3							
<b>CO5</b>	3	3	3			3						

Note: 1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High), -: No Correlation

Text Books				
Sl. No	Title of the Book	Name of the Author/s	Name of the Publisher	Edition and Year
<b>1</b>	Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition	Dan Jurafsky and James H. Martin.	Pearson	2006
<b>2</b>	Introduction to Natural Language Processing	Jacob Eisenstein	MIT Press	2019
<b>3</b>	Natural Language Processing with Transformers	Lewis Tunstall, Leandro von Werra, and Thomas Wolf	O'Reilly	2022

Reference Books				
Sl. No	Title of the Book	Name of the Author/s	Name of the Publisher	Edition and Year
1	Deep learning for Natural Language Processing	Stephan Raaijmakers	Manning	2022
2	Natural Language Processing with PyTorch	Delip Rao and Brian McMahan	O'Reilly	2019
3	Deep Learning	Ian Goodfellow, Yoshua Bengio, Aaron Courville	MIT Press	2016

Video Links (NPTEL, SWAYAM...)	
No.	Link ID
1	<a href="https://onlinecourses.nptel.ac.in/noc19_cs56">https://onlinecourses.nptel.ac.in/noc19_cs56</a>



# NATURAL LANGUAGE PROCESSING

## **PROF. PAWAN GOYAL**

Department of Computer Science and Engineering

IIT Kharagpur

**PRE-REQUISITES** : Basic knowledge of probabilities for the lectures and python for programming assignment

**INDUSTRIES APPLICABLE TO** : Microsoft Research, Google, Adobe, Xerox, Flipkart, Amazon

## **COURSE OUTLINE :**

This course starts with the basics of text processing including basic pre-processing, spelling correction, language modeling, Part-of-Speech tagging, Constituency and Dependency Parsing, Lexical Semantics, distributional Semantics and topic models. Finally, the course also covers some of the most interesting applications of text mining such as entity linking, relation extraction, text summarization, text classification, sentiment analysis and opinion mining.

## **ABOUT INSTRUCTOR :**

Prof. Pawan Goyal is an Assistant Professor at the Department of Computer Science and Engineering, IIT Kharagpur. His research interests include Natural Language Processing, Text Mining, Information Retrieval and Sanskrit Computational Linguistics. He has published around 75 research papers in international conferences and journals. He has published in various top-tier conferences and journals including ACL, NAACL, EMNLP, SIGIR, KDD, CIKM, JCDL, ICWSM, CSCW, Coling, IEEE and ACM transactions. Prior to joining IIT Kharagpur, he received his B. Tech. degree in Electrical Engineering from IIT Kanpur in 2007 and his Ph.D. degree in the faculty of Computing and Engineering from University of Ulster, UK in 2011. He was then a Post Doctoral Fellow at INRIA Paris Rocquencourt.

## **COURSE PLAN :**

**Week 1:** Introduction and Basic Text Processing

**Week 2:** Spelling Correction, Language Modeling

**Week 3:** Advanced smoothing for language modeling, POS tagging

**Week 4:** Models for Sequential tagging – MaxEnt, CRF

**Week 5:** Syntax – Constituency Parsing

**Week 6:** Dependency Parsing

**Week 7:** Distributional Semantics

**Week 8:** Lexical Semantics

**Week 9:** Topic Models

**Week 10:** Entity Linking, Information Extraction

**Week 11:** Text Summarization, Text Classification

**Week 12:** Sentiment Analysis and Opinion Mining

# SYLLABUS COMPARISON

KTU: PECST862 - Natural Language Processing  
NPTEL: Natural Language Processing

KTU SYLLABUS TOPICS	NPTEL SYLLABUS TOPICS	OK
Module 1: Text Processing, Tokenization, Morphology	Text Processing, Spelling Correction	▪
Module 2: Language Modeling, POS Tagging	Language Modeling, POS Tagging, CRF	▪
Module 3: Parsing, Syntax Analysis	Constituency Parsing, Dependency Parsing	▪
Module 4: Semantic Analysis, Word Embeddings	Distributional Semantics, Topic Models	▪
Module 5: NLP Applications, Text Classification	Entity Linking, Sentiment Analysis	▪

## CONTENT OVERLAP: $\geq 70\%$

The above comparison confirms that the NPTEL course content matches at least 70% of the KTU syllabus as required by R 17.4.