



# 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