NATURAL LANGUAGE PROCESSING

PROF. PAWAN GOYAL
Dept. of Computer Science and Engineering
IIT Kharagpur

TYPE OF COURSE : Rerun | Elective | UG/PG
COURSE DURATION : 12 weeks (29 Jul’19 - 18 Oct’19)
EXAM DATE : 16 Nov 2019

INTENDED AUDIENCE : B.E./Msc (Computer Science)
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 joined the Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur as an Assistant Professor on July 30th, 2013. Prior to that, he was working at INRIA Paris-Rocquencourt as a post doctoral fellow with Prof. Gérard Huet on The Sanskrit Heritage Site. He did his B. Tech. in Electrical Engineering from Indian Institute of Technology, Kanpur. He received his Ph. D. from Intelligent Systems Research Centre, Faculty of Computing and Engineering, University of Ulster, UK. His main research interests include Text Mining, Natural Language Processing, Information Retrieval and Sanskrit Computational Linguistics. He has published over 40 research articles in various CS journals and conferences including ACL, Coling, TKDE, CACM, KDD, CIKM, JCDL.

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: Lexical Semantics
Week 8: Distributional Semantics
Week 9: Topic Models
Week 10: Entity Linking, Information Extraction
Week 11: Text Summarization, Text Classification
Week 12: Sentiment Analysis and Opinion Mining