NATURAL LANGUAGE PROCESSING (NLP)

PMDS606L

MODULE 1: INTRODUCTION TO NLP

- Introduction to various levels (stages) of Natural Language Processing.
- Ambiguities, varieties and computational challenges in processing Natural Languages.
- Introduction to Real life applications of NLP such as spell and grammar checkers, information extraction, information retrieval, question answering, and machine translation.

MODULE 2: TEXT PROCESSING

- Text pre-processing
- Challenges
- Tokenization
- Sentence Segmentation
- Regular Expressions
- Words
- Text Normalization
- Minimum Edit Distance
- Introduction to Corpora
- Corpora Analysis.

MODULE 3: LANGUAGE MODELLING

- The role of Language Models
- N-gram Models
- Estimating Parameters and Smoothing
- Evaluating Language Models

MODULE 4: MORPHOLOGICAL ANALYSIS AND POS TAGGING

- Parts of Speech and Morphology
- Inflectional and Derivation Morphology
- Morphological Analysis
- FSA and Generation using Finite State Transducers
- Introduction to POS tagging
- HMM
- Viterbi decoding for HMM.

MODULE 5: SYNTACTIC ANALYSIS

- Ambiguities in Syntactic Parsing
- Context Free Grammar
- CYK Parsing
- Shallow Parsing and Chunking
- Dependency Parsing
- Statistical Parsing
- PCFG

MODULE 6: SEMANTIC ANALYSIS

- Semantics
- Lexical Semantics
- Word Senses
- Relations between Senses
- Word Sense Disambiguation,
- Word Similarity
- WordNet
- Thesaurus based Word Similarity
- Thematic Roles
- Semantic Role Labelling with CRFs

MODULE 7: NLTK WITH PYTHON

- Tokenizing Text and WordNet Basics
- Replacing and Correcting Words
- Part of Speech Tagging
- Extracting Chunks
- Text Classification
- Named Entity Recognition

BOOKS

1. Daniel Jurafsky and James H. Martin, Speech and Language Processing, 2017, 3rd edition, Prentice Hall..
2. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, 2016, MIT Press.
3. James Allen “Natural Language Understanding, 2012, 8th Edition, Pearson Publication.
4. Vajjala, Sowmya, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana. Practical natural language processing: A comprehensive guide to building real- world NLP systems, 2020, O'Reilly Media.

COURSE OBJECTIVES

1. To introduce the fundamental concepts and techniques of Natural language Processing for analyzing words based on Morphology and CORPUS.
2. To examine the NLP models and interpret algorithms for classification of NLP sentences by using both the traditional, symbolic and the more recent statistical approach.
3. To get acquainted with the algorithmic description of the main language levels that includes morphology, syntax, semantics, and pragmatics for information retrieval and machine translation applications.

COURSE OUTCOMES

1. Understand the fundamental concepts of natural language processing.
2. Understand the text pre-processing and corpora.
3. Analyze the words and perform POS tagging.
4. Distinguish between the syntactic and semantic correctness of the natural language.
5. Develop simple language models using NLTK

THEORY CLASSES

9 JUL	11 JUL	15 JUL	16 JUL	18 JUL	17
22 JUL	23 JUL	25 JUL	29 JUL	30 JUL	
1 AUG	2 AUG	5 AUG	6 AUG	8 AUG	
12 AUG	13 AUG				
CAT I					
26 AUG	29 AUG	30 AUG	2 SEP	3 SEP	15
9 SEP	10 SEP	12 SEP	16 SEP	17 SEP	
19 SEP	23 SEP	24 SEP	30 SEP	3 OCT	
CAT II					
14 OCT	15 OCT	17 OCT	28 OCT	29 OCT	12
31 OCT	4 NOV	5 NOV	7 NOV	11 NOV	
12 NOV	14 NOV				
FAT					

MODULE	HOURS			
Module 1: Introduction to NLP	5	CAT 1		FAT
Module 2: Text Processing	6			
Module 3: Language Modelling	6			
Module 4: Morphological Analysis and POS Tagging	7		CAT 2	
Module 5: Syntactic Analysis	6			
Module 6: Semantic Analysis	7			
Module 7: NLTK with Python	6			