Birla Institute of Technology & Science, Pilani
Hyderabad Campus

FIRST SEMESTER 2023-24
Course Handout Part II

Date: 11.08.2023

In addition to part-I (General Handout for all courses appended to the time table) this portion gives further specific details regarding the course.

Course No. : *CS F429*

Course Title : *Natural Language Processing*

Instructor-in-Charge : *Prof. Aruna Malapati*

1. Scope and Objectives

The course intends to present a fairly broad undergraduate/post-graduate level introduction to Natural Language Processing (NLP, a.k.a. computational linguistics), studying computing systems that can process, understand, or communicate in human language. The primary focus of the course will be on understanding various NLP tasks as listed on the course syllabus, algorithms for effectively solving these problems, and methods for evaluating their performance.
This subject aims to achieve the following goals:

- To introduce students to the challenges of empirical methods for natural language processing (NLP) applications.
- To introduce basic mathematical models and methods used in NLP applications to formulate computational solutions.
- To provide students with knowledge on designing procedures for natural language resource annotation and using related tools for text analysis and hands-on experience of using such tools.
- To introduce students to research and development work in information retrieval, information extraction, and knowledge discovery using different natural language resources.
- To give an overview of the major technologies in speech recognition and synthesis including tools for acoustic analysis and hands-on experience of using such tools
- To give students opportunities to sharpen their programming skills for computational linguistics applications

Note: Programming in Java or C, however, programming in Python will be an advantage, and knowledge of core data structures and algorithms.

2.a. Text Book

- **T1:** Jurafsky and Martin, SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Third Edition, McGraw Hill, 2008.

b. Reference Books

- **R1:** Uday Kamath, John Liu, James Whitaker, Springer, Deep Learning for NLP and Speech Recognition, Springer, 2019.
- **R2:** Uday Kamath Kenneth L. Graham Wael Emara , Transformers for Machine Learning A Deep Dive, CRC Press, Tylor and Francis group

- **R3**: Natural Language Toolkit. Bird and Loper, and other developers. Available for free at: – http://www.nltk.org/

3. Course Plan

Lecture No	Learning Objectives	Topics to be covered	Reference
1,2	• To Introduce NLP and its applications	Introduction to NLP, Model of natural language	T1:Ch1
3	● To apply N-gram models for document generation	N-Gram Language Models and their evaluation	T1:Ch3
4	● List NLP applications	POS tagging, Parsing, Named-Entity Recognition, Semantic Role Labeling, Sentiment Classification, Machine Translation, Question Answering, Dialogue Systems	T1: Basic Introduction to each task Ch 8,13,20,11,25, 26 R2 Ch3
5	● To convert words into various forms of vectors	Vector semantics and Embeddings: TF-IDF, Pointwise Mutual Information	T1:Ch6 Class Notes
6	● Design and train neural networks to generate pre-trained word embeddings	Introduction to Neural Networks and pre-trained word embeddings	R2 Ch 4.1-4.5
7-9	● Design a neural network to pre-train word vectors using Wikipedia	Skipgram, CBOW, Glove	R2 Ch5 Class Notes
10-26	• Design different neural networks for capturing the context and meaning of the words.	RNN, LSTM, ELMO, Attention and transformers, BERT variants, GPT, Multilingual Language Models	R2 Ch2,Ch3,Ch4 R2 Ch5,9
27-28	• Identify the problems with word embeddings and evaluate the importance of sub word embeddings	Beyond Word Embeddings: Subword Embeddings, Word Vector Quantization, Sentence Embeddings and others	R1 Ch 5.3,5.4
29	● To apply Machine Translation on a given parallel corpus and measure the performance of the translation	Machine Translation and evaluation using BLUE scores	T1:Ch11
30	• To evaluate and measure the performance of models	Performance evaluation of different models on NLP tasks	Class Notes
31-35	● To evaluate the	Transfer and Multitask Learning in NLP	R1 Ch 10

	performance of the primary task by jointly training two models		
36-37	● To apply and explain the document generation process using LDA	Topic Modelling using Latent Dirichlet allocation	Class Notes
38-40	• To apply HMMs for POS tagging	Introduction to Hidden Markov Models (HMMs) for POS	T1:Ch8

4. Evaluation Scheme

Component	Duration	Weightage	Date & Time	Nature of Component
Project/Assignments	TBA	**20%**	--	Open Book
Class Participation*	During class hours	**10%**		Open Book
Mid-Term exam	90 mins	**30%**	09/10 - 11.30 - 1.00PM	Closed Book
Comprehensive exam	3 hours	**40%**	06/12 AN	Closed Book

***Class Participation: The students will be asked to work on a task specific on the content covered. The best n-3 submissions will be considered out of n.**
Note: minimum 40% of the evaluation will be completed by Midsem grading.

5. Chamber Consultation: Mon 4-5 Pm @ H132

6. Notices: CMS

7. Make-up Policy: There is no makeup for class participation and Project/Assignments. The I/C shall grant make-ups for Mid sem tests on prior permission and only to genuine cases. Make-up for the comprehensive examination will be decided by the I/C only for genuine cases and scheduled by the AUGSD.

8. Academic Honesty and Integrity Policy: Academic honesty and integrity are to be maintained by all the students throughout the semester and no type of academic dishonesty is acceptable.

Instructor-in-charge