



FIRST SEMESTER 2022-23
Course Handout Part II

Date: 29.08.2022

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:** Manning and Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.
- **R2:** Uday Kamath, John Liu, James Whitaker, Springer, Deep Learning for NLP and Speech Recognition, Springer, 2019.

- **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	Chapter in the TextBook
1	● To Introduce NLP and its applications	Introduction to NLP	T1:Ch1
2	●To apply N-gram models for document generation	N-Gram Language Models and their evaluation	T1:Ch3
3	●To convert words into various forms of vectors	Vector semantics and Embeddings: TF-IDF, Pointwise Mutual Information	T1:Ch6 Class Notes
4-6	●Design and train neural networks to generate pre-trained word embeddings	Introduction to Neural Networks and Pretrained word embeddings	R2 Ch4.1-4.5 R2 Ch5
7-8	• To apply POS tagging on a given corpus	Part-of-Speech Tagging, Sequence Labeling, using Hidden Markov Models (HMMs)	T1:Ch8
9-12	●To apply and explain the document generation process using LDA	Topic Modelling using Latent Dirichlet allocation	Class Notes
13-19	●To apply Statistical Machine Translation on a given parallel corpus and measure the performance of the translation	Statistical Machine Translation: Word and Phrase-based Models, BLUE scores	T1:Ch11
20-23	●Evaluate the use of context-free grammar for parsing	Constituency Grammars	T1:Ch12
24-26		Statistical and dependency parsing	T1:Ch14&15
27-29	●To extract the semantic meaning of sentences	Logical Representations of Sentence Meaning	T1:Ch16
30-35	●To apply information extraction techniques for various application	Information Extraction: Named Entity Recognition, Relation extraction, Extracting Events and Time	T1:Ch18
35-38	• To apply Word-sense disambiguation	Word Senses and WordNet	T1:Ch19
40-42	●To apply all the NLP-related techniques for solving Question Answering and	Question Answering, Dialog Systems and Chatbots	T1:Ch25,26

	Summarization		
--	---------------	--	--

4. Evaluation Scheme

Component	Duration	Weightage	Date & Time	Nature of Component
Project Phase1 evaluation – 10% before mid sem Phase2 evaluation – 15% before compre	Take home	25%	--	Take Home Open Book
Mid-Term exam	90 mins	35%	2/11/2022 9.00 – 10.30 AM	Closed Book
Comprehensive exam	3 hours	40%	22/12/202 2 FN	Closed Book

Note: minimum 40% of the evaluation to be completed by midsem grading.

5. Chamber Consultation: TBA

6. Notices: CMS

7. Make-up Policy: Make-ups for Mid sem tests shall be granted by the I/C on prior permission and only to genuine cases. Make-up for the comprehensive examination will be decided by the I/C 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