Birla Institute of Technology & Science, Pilani, K. K. Birla Goa Campus
First Semester 2021 - 22

In addition to the Part-I (General Handout for all courses appended to the timetable) this portion gives further specific details on course.
Course ID: **CS F429**, Course Title: **Natural Language Processing (NLP)**
Class Timings: **Mon, Wed, Fri 11:00 - 11:50 AM**, (Extra: Oct 12, Nov 18)
Instructor-In-charge: **Dr. Swati Agarwal** (swatia@goa.bits-pilani.ac.in)

1. **Prerequisites:** Intermediate programming skills (preferably Python) and Data Structures & Algorithms.
2. **Course Objectives and Scope**
   The intent of the course is to present a fairly broad FD level introduction to NLP. The primary focus of the course will be on understanding various NLP tasks, algorithms for effectively solving these problems, and methods for evaluating their performance.
3. **Text Books**

   T1. Dan Jurafsky, James H. Martin. "*Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*". Prentice-Hall. 3^{rd} edition [Official Link for 3^{rd} edition draft.]
4. **Reference Books**

   R1. Manning, Christopher, and Hinrich Schutze. "*Foundations of statistical natural language processing*". MIT press, 1999.

   R2. Steven Bird, Ewan Klein, and Edward Loper. "*Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*". O'Reilly.

   R3. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. 2020. "*Practical Natural Language Processing*". O'Reilly.

   R4. Hobson Lane, Cole Howard, Hannes Hapke. 2019. "*Natural Language Processing in Action*". Manning Publications [Live Book]
5. **Course Plan:** Some part of the course (pre and post midsem) will be conducted in flipped mode where students will be provided the video lectures and the doubts clearing sessions will be conducted in live classes. The course does not have a lab component; however, tutorials will be conducted in class for hands-on and project guidance.

#Lecture	Module	Topics	Reference
1	-	Overview	course handout
2-3	I	Introduction, Basic Text Processing: Tokenization, Stemming	T1 Ch2
4	II	Language Modeling: N-grams, smoothing	T1 Ch2, Ch3
3	III	Morphology, Parts of Speech Tagging	T1 Ch8
7	IV	Syntax: CFGs, Dependency Parsing	T1 Ch12, 13, 14
4	V	Topic Models	R4 Ch4
3	VI	Distributional Semantics	Research Papers
4	VII	Lexical Semantics, Word Sense Disambiguation	T1 Ch18
4	VIII	Information Extraction: Relation extraction	T1 Ch17
3	IX	NLP Applications. Q&A, Summariation, Sentiment Analysis	Class Notes

6. **Evaluation Scheme**

S.No.	Component	Weightage	Date	Time	Remarks
1	Mid-sem Exam	35%	Oct 21	9:00-10:30 AM	Closed Book
2	Project	25%	TBA	TBA	NA
3	End-sem Exam	40%	Dec 22	3:00-5:00 PM	Closed Book

7. **Notice**: All notices concerning this course will be displayed on the Google Classroom. Keep an eye on ID/ARC notices as well.
8. **Malpractice Regulations:** Any attempt of cheating or plagiarism in midsem or endsem will attract disciplinary committee action. Any student involved in malpractice in project will be awarded negative marks equal to the weightage of that component.
9. **Make-up Policy**
   - Make-up shall be granted only in genuine cases based on individual's need and circumstances and must be approved by the ID.
   - No marks will be awarded without make-up for that component.
   - No make-up for project component.

Instructor In-charge
Dr. Swati Agarwal