

| Course Code | Course Title | | | | Core/ Elective | | |
|--------------------------|-----------------------------|---|---|---|----------------|-----|---------|
| U21CD801 | NATURAL LANGUAGE PROCESSING | | | | ELECTIVE | | |
| Prerequisite | Contact Hours Per Week | | | | CIE | SEE | Credits |
| | L | T | D | P | | | |
| Python, Machine Learning | 3 | - | - | - | 40 | 60 | 3 |

Course Objectives

This course will introduce students to:

1. Teach students the leading trends and systems in natural language processing.
2. Make them understand the concepts of morphology, syntax and semantics of the language and that they are able to give the appropriate examples that will illustrate the above mentioned concepts.
3. Teach them to recognize the significance of pragmatics for natural language understanding.
4. Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic and semantic processing.
5. How to evaluate the strengths and weaknesses of various NLP technologies and frameworks as they gain practical experience in the NLP toolkits available.

Course Outcomes

1. To tag a given text with basic Language features
2. To design an innovative application using NLP components
3. To implement a rule based system to tackle morphology/syntax of a language
4. To design a tag set to be used for statistical processing for real-time applications
5. To compare and contrast the use of different statistical approaches for different types of NLP applications and Perform various language phonetic analysis

UNIT I

Introduction of NLP: Origins and challenges of NLP, Language Modeling: Grammar-based LM, Statistical LM – Regular Expressions, Automata , Morphology and Finite State Transducers, Tokenization, stemming, Normalization, Detecting and Correcting Spelling Errors, Minimum Edit Distance.

UNIT II

WORD LEVEL ANALYSIS: N-grams, Evaluating N-grams, Smoothing, Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Entropy, Hidden Markov and Maximum Entropy models, ; Named Entities

UNIT-III

SYNTACTIC ANALYSIS: Context free rules and trees – The noun Phrase – Co-ordination – Verb phrase – context free grammars – Parsing with context free grammars, Shallow parsing – Probabilistic CFG , Dependency Grammar , Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion- Reference Resolution- Discourse Coherence and Structure.

UNIT_IV

Speech Fundamentals: Phonetics – speech sounds and phonetic transcription – articulatory phonetics – phonological categories and pronunciation variation – acoustic phonetics and signals – phonetic resources – articulatory and gestural phonology

UNIT-V

Speech synthesis – text normalization – phonetic analysis – prosodic analysis – diphone waveform synthesis – unit selection waveform synthesis – evaluation

Suggested reading:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
2. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, First Edition, OReilly Media, 2009.