MCPE2005 NATURAL LANGUAGE PROCESSING (3-0-0)

COURSE OBJECTIVES:

1. To teach the fundamentals of NLP, and also to make them for understanding CFG, PCFG in NLP.
2. To know the role of semantics of sentences and pragmatic.
3. To teach the basic concepts of speech processing along with analysis and modeling.

MODULE-1

Introduction: Origins and challenges of NLP, Language modeling: grammar-based LM, statistical LM regular expressions, finite-state automata, English morphology, transducers for lexicon and rules, tokenization, detecting and correcting spelling errors, minimum edit distance.

Word Level Analysis: Unsmoothed n-grams, evaluating n-grams, smoothing, interpolation and backoff – word classes, part-of-speech tagging, rule-based, stochastic and transformation-based tagging, issues in pos tagging, hidden markov and maximum entropy models.

MODULE-2

Syntactic Analysis: Context free grammars, grammar rules for english, treebanks, normal forms for grammar, dependency grammar, syntactic parsing, ambiguity, dynamic programming parsing, shallow parsing, probabilistic CFG, probabilistic CYK, probabilistic lexicalized CFGs, feature structures, unification of feature structures**.**

MODULE-3

Semantics and Pragmatics: Requirements for representation, first-order logic, description logics, syntax-driven semantic analysis, semantic attachments, word senses, relations between senses, thematic roles, selectional restrictions, Word Sense Disambiguation(WSD), WSD using supervised, dictionary & thesaurus, bootstrapping methods, word similarity using thesaurus and distributional methods.

Speech fundamentals: articulatory phonetics, production and classification of speech sounds; acoustic phonetics, acoustics of speech production, review of digital signal processing concepts, short-time fourier transform, filter- bank and LPC methods.

MODULE-4

Speech-Analysis and Speech Modeling: Features, feature extraction and pattern comparison techniques: speech distortion measures, mathematical and perceptual, log spectral distance, cepstral distances, weighted cepstral distances and filtering, likelihood distortions, spectral distortion using a warped frequency scale, lpc, plp and mfcc coefficients, time alignment and normalization, dynamic time warping, multiple time, alignment paths. Hidden markov models: markov processes, HMMS – evaluation, optimal state sequence, viterbi search, baum-welch parameter re-estimation, implementation issues.

COURSE OUTCOMES:

CO 1. Learn the fundamentals of NLP andUnderstand the use of CFG and PCFG in NLP

CO 2. Understand the role of semantics of sentences and pragmatic

CO 3. Introduce Speech Production and Related Parameters of Speech.

CO 4. Show the Computation and use of Techniques such as Short Time Fourier Transform,Linear Predictive Coefficients and other Coefficients in the Analysis of Speech.

REFERENCE BOOKS:

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.
3. Lawrence Rabiner And Biing-Hwang Juang, "Fundamentals Of Speech Recognition", Pearson Education, 2003.
4. Daniel Jurafsky and James H Martin, "Speech And Language Processing – An Introduction To Natural Language Processing, Computational Linguistics, And Speech Recognition", Pearson Education, 2002