



Academic Year: 2024-2025

Mid Term Test 1 Syllabus

Course Name: Natural Language Processing

Class: Final Year (C1, C2, C3)

Course Code: DJ19CEEC8011

Sem: VIII

Unit No	Topics
1	Introduction: History of NLP, Generic NLP System, Levels of NLP, Knowledge in Language Processing, Ambiguity in Natural Language, Stages in NLP, Challenges of NLP, Applications of NLP
2	Word Level Analysis: Morphology Analysis –Survey of English Morphology, Inflectional Morphology & Derivational Morphology, Lemmatization, Regular Expression, Finite Automata, Finite State Transducers (FST), Morphological Parsing with FST, Lexicon Free FST, Porter Stemmer. N –Grams, Unigrams/Bigrams Language Models, Corpora, Computing the Probability of Word Sequence, Training and Testing. Perplexity and Entropy: Smoothing and Backup, Zipf's Law, Add One Smoothing, Witten-Bell Discounting, Good Turing Discounting, Back Off Methods, Class Based Models, Google N-Gram Release.
3	Syntax Analysis: Part-Of-Speech Tagging (POS) - Open and Closed Words. Tag Set for English (Penn Treebank), Rule Based POS Tagging, Transformation Based Tagging, Stochastic POS Tagging and Issues –Multiple Tags & Words, Unknown Words. Hidden Markov Model (HMM), Maximum Entropy, And Conditional Random Field (CRF). CFG: Derivations, Constituency, Phrase Structure and Dependency Structure

Subject Incharge,

Dr. Kiran Bhowmick

Dr. Chetashri Bhadane