# 20AD32E1 - NATURAL LANGUAGE PROCESSING

|  |  |  |  |
| --- | --- | --- | --- |
| Course Category: | Professional Elective | Credits: | 3 |
| Course Type: | Theory | Lecture-Tutorial-Practical: | 3-0-0 |
| Prerequisite: | Knowledge of basic Machine learning concepts.  Knowledge of mathematical concepts such as Probability, Linear Algebra and Calculus is must. | Sessional Evaluation:  Univ. Exam Evaluation:  Total Marks: | 40  60  100 |
| Objectives: | * Learn the concepts of Natural Language processing. * Gain practical understanding of relevant terminology, concepts in Natural Language Processing. | | |

|  |  |  |
| --- | --- | --- |
| Course Outcomes | Upon successful completion of the course, the students will be able to: | |
| CO1 | Explain the fundamental models of Natural Language Processing systems |
| CO2 | Design Finite-State Transducers for English Morphology. |
| CO3 | Take part in word prediction of Language sentences. |
| CO4 | Interpret the computational models for assigning POS tagging. |
| CO5 | Apply basic Top-Down Parser for syntax analysis of Natural Language sentences. |
| CO6 | Describe probabilistic grammars and parsers for syntax analysis of Natural Language sentences. |
| Course Content | UNIT-I  **Introduction:**  Knowledge in Speech and Language Processing, Ambiguity, Models and Algorithms, Language, Thought and Understanding, The State of the Art, Some Brief History.  UNIT-II  **Words and Transducers:**  Survey of English Morphology, Finite-State Morphological Parsing, Construction of a finite state lexicon, Finite-State Transducers, FSTs for Morphological Parsing, Transducers and Orthographic rules, The Combination of an FST Lexicon and Rules.  UNIT-III  **N-grams:**  Word counting in Corpora, Simple (Unsmoothed) N-grams, Training and Test Sets, Evaluating N-GRAMS, Smoothing- Laplace smoothing and Good-Turing Discounting, Interpolation, Back-off.  UNIT-IV  **Part-of-Speech Tagging:**  English word classes, Tagsets for English, Part-of-speech Tagging, Rule-Based Part-of-Speech Tagging, HMM Part- of-Speech tagging- Computing the most-likely tag sequence, Formalizing Hidden Markov Model taggers, The Viterbi Algorithm for HMM Tagging.  UNIT-V  **Syntactic Parsing:**  Parsing as Search- Top Down Parsing, Bottom-Up Parsing, Ambiguity, Search in the face of ambiguity, Dynamic programming parsing methods- CKY parsing, The Earley algorithm, Chart Parsing.  UNIT-VI  **Statistical Parsing:**  Probabilistic Context-Free Grammars for Disambiguation, Probabilistic Context-Free Grammars for Language Modeling, Probabilistic CKY Parsing of PCFGs, Learning PCFG Rule Probabilities, Problems with PCFGs, Improving PCFGs by Splitting and Merging Nonterminals. | |
| Text Books &  Reference  Books | **TEXT BOOKS:**   1. An introduction to natural language processing, computational linguistics, and speech recognition. Daniel Jurafsky & James H. Martin.   **REFERENCE BOOKS:**   1. Christopher D. Manning, Hinrich Schutze, Foundations of Statistical Natural Language Processing, (1999), The MIT Press. 2. Tanveer Siddiqui, US Tiwary, Natural Language Processing and Information Retrieval, (2008), Oxford University Press. | |
| E-Resources | 1. <https://nlp.stanford.edu/fsnlp/> | |