

PCS-210 NATURAL LANGUAGE PROCESSING

L T P Cr

3 0 2 4.0

Course Objectives: To understand the advanced concepts of Natural Language Processing and to be able to apply the various concepts of NLP in other application areas.

Introduction: Origin of Natural Language Processing (NLP), Challenges of NLP, NLP Applications, Processing Indian Languages.

Words and Word Forms: Morphology fundamentals; Morphological Diversity of Indian Languages; Morphology Paradigms; Finite State Machine Based Morphology; Automatic Morphology Learning; Shallow Parsing; Named Entities; Maximum Entropy Models; Random Fields, Scope Ambiguity and Attachment Ambiguity resolution.

Machine Translation: Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation, UNL Based Machine Translation, Translation involving Indian Languages.

Meaning: Lexical Knowledge Networks, WorldNet Theory; Indian Language WordNets and Multilingual Dictionaries; Semantic Roles; Word Sense Disambiguation; WSD and Multilinguality; Metaphors.

Speech Recognition: Signal processing and analysis method, Articulation and acoustics, Phonology and phonetic transcription, Word Boundary Detection; Argmax based computations; HMM and Speech Recognition.

Other Applications: Sentiment Analysis; Text Entailment; Question Answering in Multilingual Setting; NLP in Information Retrieval, Cross-Lingual IR.

Laboratory Work: To implement Natural language concepts and computational linguistics concepts using popular tools and technologies. To implement key algorithms used in Natural Language Processing.

Recommended Books:

1. Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press (2008).
2. Allen J., Natural Language understanding, Benjamin/Cummings, (1987).
3. Jensen K., Heidorn G.E., Richardson S.D., Natural Language Processing: The PLNLP Approach, Springer (2013).
4. Roach P., Phonetics, Oxford University Press (2012).