

DSPE3006 NATURAL LANGUAGE PROCESSING (3-0-0)

Course objectives:

1. Introduce to some of the problems and solutions of NLP and their relation to linguistics and statistics.
2. To understand linguistic phenomena and learn to model them with formal grammars.
3. To understand and carry out proper experimental methodology for training and evaluating empirical NLP systems.
4. To learn how to manipulate probabilities, construct statistical models over string and trees
5. To estimate parameters using supervised and unsupervised training methods.
6. To design, implement, and analyse NLP algorithms. Able to design different language modeling Techniques.

Module – I (08 Hours)

Introduction to NLP:

Applications or Use cases of NLP, Components of NLP, Steps in NLP, Finding the Structure of Words: Words and Their Components, Lexemes, Morphemes, Morphology, Problems in morphological processing, Typology, Morphological Typology.

Text Preprocessing (NLP with python NLTK package): Word Tokenization, Sentence Tokenization, Filtering Stop words, Stemming, Tagging Parts of Speech, Lemmatization, Chunking, Chinking, Named Entity Recognition, Term Frequency and Inverse Document Frequency (TF-IDF).

Module – II (08 Hours)

Syntax Analysis:

Parsing Natural Language, Tree banks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure: Syntax Analysis using Dependency Graph, Syntax Analysis using Phrase Structure Trees, Parsing Algorithms: Shift Reduce Parsing, Hyper Graphs and Chart Parsing (CYK Parsing), Models for ambiguity Resolution in Parsing: Probabilistic Context Free Grammar, Generative Models, Discriminative models for Parsing.

Module –III (08 Hours)

Language Modeling:

Introduction, N-Gram Models, Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems.

Module – IV (08 Hours)

Semantic Parsing:

Introduction, Semantic Interpretation, System Paradigms, Word Sense Systems.

Predicate-Argument Structure: Meaning Representation Systems. Discourse Processing: Cohesion, Reference Resolution, Discourse Cohesion and Structure.

Module – V (08 Hours)

Information retrieval and lexical resources:

Information Retrieval: Design features of Information Retrieval Systems-Classical, Non classical, Alternative Models of Information Retrieval – valuation Lexical Resources: World Net-Frame Net- Stemmers-POS Tagger-Research Corpora.

Course outcomes:

- Analyse the natural language text.
- Generate the natural language.
- Demonstrate Text mining.
- Apply information retrieval techniques.
- Show sensitivity to linguistic phenomena and an ability to model them with formal grammars.
- Understand and carry out proper experimental methodology for training and evaluating empirical NLP systems.
- Able to manipulate probabilities, construct statistical models over strings and trees.
- Able to design, implement, and analyze NLP algorithms. Able to design different language modeling Techniques.

Text Books:

1. Multilingual natural Language Processing Applications: From Theory to Practice–Daniel M.Bikel and Imed Zitouni, Pearson Publication.
2. Speech and Natural Language Processing-Daniel Jurafsky & James H Martin, Pearson Publications.

Reference books:

1. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S.Tiway.