

## **CS8084 NATURAL LANGUAGE PROCESSING**

### DETAILED SYLLABUS

#### **OBJECTIVES:**

- To learn the fundamentals of natural language processing
- To understand the use of CFG and PCFG in NLP
- To understand the role of semantics of sentences and pragmatics
- To apply the NLP techniques to IR applications

#### **UNIT I 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

#### **UNIT II WORD LEVEL ANALYSIS**

Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Back off – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.

#### **UNIT III 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.

#### **UNIT IV 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 using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods.

#### **UNIT V DISCOURSE ANALYSIS AND LEXICAL RESOURCES**

Discourse segmentation, Coherence – Reference Phenomena, Anaphora Resolution using Hobbs and Centering Algorithm – Coreference Resolution – Resources: Porter Stemmer, Lemmatizer, Penn Treebank, Brill's Tagger, WordNet, PropBank, FrameNet, Brown Corpus, British National Corpus (BNC).

#### **TEXT 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 PythonII, First Edition, O'Reilly Media, 2009.

**REFERENCES:**

1. Breck Baldwin, —Language Processing with Java and Ling Pipe Cookbook, Atlantic Publisher, 2015.
2. Richard M Reese, —Natural Language Processing with Javall, OReilly Media, 2015.
3. Nitin Indurkha and Fred J. Damerau, —Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
4. Tanveer Siddiqui, U.S. Tiwary, —Natural Language Processing and Information Retrieval, Oxford University Press, 2008.