Natural Language Processing
CS801C
Contracts: 3L
Credits- 3

Module I
Regular Expressions and Automata (Recap) [**2L**]
Introduction to NLP, Regular Expression, Finite State Automata
Tokenization [**5L**]
Word Tokenization, Normalization, Sentence Segmentation, Named Entity Recognition,
Multi Word Extraction, Spell Checking - Bayesian Approach, Minimum Edit Distance
Morphology [**4L**]
Morphology - Inflectional and Derivational Morphology, Finite State Morphological Parsing, The Lexicon
and Morphotactics, Morphological Parsing with Finite State Transducers, Orthographic Rules and Finite
State Transducers, Porter Stemmer

Module II
Language Modeling [**4L**]
Introduction to N-grams, Chain Rule, Smoothing - Add-One Smoothing, Witten-Bell Discounting; Backoff,
Deleted Interpolation, N-grams for Spelling and Word Prediction, Evaluation of language models.
Hidden Markov Models and POS Tagging [**4L**]
Markov Chain, Hidden Markov Models, Forward Algorithm, Viterbi Algorithm, Part of Speech Tagging -
Rule based and Machine Learning based approaches, Evaluation

Module III
Text Classification [**4L**]
Text Classification, Naïve Bayes' Text Classification, Evaluation, Sentiment Analysis - Opinion Mining and
Emotion Analysis, Resources and Techniques
Context Free Grammar [**5L**]
Context Free Grammar and Constituency, Some common CFG phenomena for English, Top-Down and
Bottom-up parsing, Probabilistic Context Free Grammar, Dependency Parsing

Module IV
Computational Lexical Semantics [**4L**]
Introduction to Lexical Semantics - Homonymy, Polysemy, Synonymy, Thesaurus - WordNet,
Computational Lexical Semantics - Thesaurus based and Distributional Word Similarity
Information Retrieval [**5L**]
Boolean Retrieval, Term-document incidence, The Inverted Index, Query Optimization, Phrase Queries,
Ranked Retrieval - Term Frequency - Inverse Document Frequency based ranking, Zone Indexing, Query
term proximity, Cosine ranking, Combining different features for ranking, Search Engine Evaluation,
Relevance Feedback

Books:
1. Speech and Language Processing, Jurafsky and Martin, Pearson Education
2. Foundation of Statistical Natural Language Processing, Manning and Schutze, MIT Press