NATURAL LANGUAGE PROCESSING
(ELECTIVE-III)

Course Code: 13CS1108
L T P C
4 0 0 3

Course Educational Objectives:
To lay out the mathematical and linguistic foundations for Natural Language Processing.

Course Outcomes:
❖ To introduce statistical methods and models to process natural languages.
❖ To introduce n-gram models, markov models

UNIT-I (12 Lectures)
INTRODUCTION:
Rationalist and Empiricist Approaches to Language, Scientific Content, Questions that linguistics should answer, Non-categorical phenomena in language, Language and cognition as probabilistic phenomena, The Ambiguity of Language: Why NLP Is Difficult, Dirty Hands, Lexical resources, Word counts, Zipf’s laws, Collocations, Concordances.

UNIT-II (12 Lectures)
MATHEMATICAL FOUNDATIONS:
Elementary Probability Theory, Probability spaces, Conditional probability and independence, Bayes’ theorem, Random variables, Expectation and variance, Notation, Joint and conditional distributions, Determining, Standard distributions, Bayesian statistics, Essential Information Theory, Entropy, Joint entropy and conditional entropy, Mutual information, The noisy channel model, Relative entropy or Kullback-Leibler divergence, The relation to language:
Cross entropy, The entropy of English, Perplexity.

UNIT-III (12 Lectures)

LINGUISTIC ESSENTIALS:
Parts of Speech and Morphology, Nouns pronouns, Words that accompany nouns: Determiners and adjectives, Verbs, Other parts of speech, Phrase Structure, Phrase structures, Dependency: Arguments and adjuncts, X’ theory, Phrase structure ambiguity, Semantics and pragmatics.

WORDS COLLOCATIONS:

UNIT-IV (12 Lectures)

STATISTICAL INFERENCE: N-GRAM MODELS OVERSPARSE DATA
BINS:
Forming Equivalence Classes, Reliability vs. discrimination, n-gram models, Statistical Estimators, Maximum Likelihood Estimation, Laplace’s law, Lidstone’s law and the Jeffreys-Perks law, Held out estimation, Cross-validation (deleted estimation), Good-Turing estimation, Briefly noted, Combining Estimators, Simple linear interpolation, Katz’s backing-off, General linear interpolation, Briefly noted Language models for Austen.

UNIT-V (12 Lectures)

MARKOV MODELS:
TEXT BOOK:

REFERENCES:

WEB REFERENCE:
https://www.coursera.org/course/nlp