19CSE453	NATURAL LANGUAGE PROCESSING	L-T-P-C: 2-0-3-3

Pre-Requisite(s):19MAT111 Multivariable Calculus, 19MAT112 Linear Algebra, 19MAT205 Probability and Random processes

Course Objectives

- This course is devoted to the study of phonological, morphological and syntactic processing. These areas will be approached from both a linguistic and an algorithmic perspective.
- The course will focus on the computational properties of natural languages and of the algorithms used to process them, as well as the match between grammar formalisms and the linguistic data that needs to be covered.

Course Outcomes

CO1: Understand the models, methods, and algorithms of statistical Natural Language Processing (NLP) for common NLP tasks.
CO2: Understand mathematical and statistical models for NLP.
CO3: Understand linguistic phenomena and linguistic features relevant to each NLP task.
CO4: Apply probabilistic models in code.
CO5: Apply learning models to NLP tasks such as speech recognition, machine translation, spam filtering, text classification, and spell checking

CO-PO Mapping

PO/PSO / CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2	2	3									3	2
CO2	3	2	3	2									3	2
CO3	3	2	3	2									3	2
CO4	3	1	2	2	3								3	2
CO5	3	1	2	2	3								3	2

Syllabus

Unit 1
Introduction- Human languages, models, ambiguity, processing paradigms; Phases in natural language processing, applications. Text representation in computers, encoding schemes. Linguistics resources- Introduction to corpus, elements in balanced corpus, TreeBank, PropBank, WordNet, VerbNet etc. Resource management with XML, Management of linguistic data with the help of GATE, NLTK. Regular expressions, Finite State Automata, word recognition, lexicon. Morphology, acquisition models, Finite State Transducer, N-grams, smoothing, entropy, HMM, ME, SVM, CRF.

Unit 2

Part of Speech tagging- Stochastic POS tagging, HMM, Transformation based tagging (TBL), Handling of unknown words, named entities, multi word expressions. A survey on natural language grammars, lexeme, phonemes, phrases and idioms, word order, agreement, tense, aspect and mood and agreement, Context Free Grammar, spoken language syntax. Parsing- Unification, probabilistic parsing, TreeBank. Semantics- Meaning representation, semantic analysis, lexical semantics, WordNet Word Sense Disambiguation- Selectional restriction, machine learning approaches, dictionary based approaches.

Unit 3

Discourse- Reference resolution, constraints on co-reference, algorithm for pronoun resolution, text coherence, discourse structure. Applications of NLP- Spell-checking, Summarization Information Retrieval- Vector space model, term weighting, homonymy, polysemy, synonymy, improving user queries. Machine Translation–EM algorithm - Discriminative learning - Deep representation learning - Generative learning.

Text Book(s)

Martin JH, Jurafsky D. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River: Pearson/Prentice Hall; 2009.

Reference(s)

James A.. Natural language Understanding,Second Edition, Pearson Education;1994.
Bharati A., Sangal R., Chaitanya V.. Natural language processing: a Paninian perspective, PHI; 2000.
Tiwary U S, Siddiqui T. Natural language processing and information retrieval. Oxford University Press, Inc.; 2008.

Evaluation Pattern

Assessment	Internal	External
Periodical 1	10	
Periodical 2	10	
*Continuous Assessment (Theory) (CAT)	10	
Continuous Assessment (Lab) (CAL)	40	
End Semester		30

*CA – Can be Quizzes, Assignment, Projects, and Reports.