NATURAL LANGUAGE PROCESSING

Course Code	23CS4602C	Year	III	Semester	II
Course Category	Professional Elective-III	Branch	CSE	Course Type	Theory
Credits	3	L – T – P	3-0-0	Prerequisites	Theory of Computation, Compiler Design, Machine Learning, Probabilistic
Continuous Evaluation:	30	Semester End Evaluation:	70	Total Marks:	100

Course Outcomes		
Upon successful completion of the course, the student will be able to:		
CO1	Understand the fundamental concepts of Natural language processing.	L2
CO2	Apply N-gram models, PoS tagging, and syntactic parsing with probabilistic grammars for comprehensive word-level and structural language analysis.	L3
CO3	Apply logic-based representations and word sense disambiguation techniques to analyze meaning and context in natural language.	L3
CO4	Apply computational techniques and lexical resources for effective discourse and reference resolution.	L3

Syllabus

Unit No.	CONTENTS	Mapped CO
I	**INTRODUCTION:** Origins and challenges of NLP – Language Modeling: Grammar-based LM, Statistical LM — English Morphology, Tokenization, Detecting and Correcting Spelling Errors. **LANGUAGE MODELING:** Introduction, N-Gram Models , Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems, Multilingual and Cross lingual Language Modeling	CO1,CO2
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.	CO1,CO2

III	**SYNTACTIC ANALYSIS**: 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	**CO1,CO2**
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.	**CO1,CO3**
V	**DISCOURSE ANALYSIS AND LEXICAL RESOURCES:** Discourse segmentation, Coherence – Reference Phenomena, Discourse Cohension, 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). Real world NLP Case Studies.	**CO1,CO4**

Learning Resources

Text Books

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, 2nd Edition, Daniel Jurafsky, James H. Martin - Pearson Publication,2014.
2. Natural Language Processing with Python, First Edition, Steven Bird, Ewan Klein and Edward Loper, OReilly Media,2009.
3. Multilingual natural Language Processing Applications: From Theory to Practice – Daniel M. Bikel and Imed Zitouni, Pearson Publication

Reference Books

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

E-Resources & other digital material

1. https://web.stanford.edu/~jurafsky/slp3/
2. https://swayam.gov.in/nd1_noc19_cs56/preview
3. https://online.stanford.edu/courses/xcs224n-natural-language-processing-deep-learning
4. https://www.coursera.org/specializations/natural-language-processing