Semester	VI	L	T	P	C	COURSE CODE
Regulation	V23	3	0	0	3	V23CSTPE10
Name of the Course	Natural Language Processing (Professional Elective-III)					
Branch	Common to CSE and CST					

Syllabus Details

Course Outcomes: After Successful completion of the Course, the student will be able to:

CO1: Explain the fundamental concepts of language modeling, English morphology, and spelling correction techniques using regular expressions and finite-state automata. **(K2)**

CO2: Apply N-gram models and part-of-speech tagging techniques using stochastic and rule-based approaches to perform word-level language analysis. **(K3)**

CO3: Construct syntactic parsers using context-free grammars and apply probabilistic parsing methods such as PCFG and CYK for syntactic analysis. **(K3)**

CO4: Describe semantic representation techniques and methods for word sense disambiguation using supervise and knowledge-based approaches. **(K2)**

CO5: Apply discourse segmentation and reference resolution techniques using lexical resources like WordNet, Penn Treebank, and Brill's Tagger for effective text analysis. **(K3)**

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 Backoff – 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).

Textbooks:

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.

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.