RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Artificial Intelligence and Machine Learning/ Artificial Intelligence and Machine Learning, V-Semester

Open Elective AL 504 (B) Natural Language Processing

COURSE OBJECTIVES: Students should develop a basic understanding in natural language processing methods and strategies and to evaluate the strengths and weaknesses of various Natural Language Processing (NLP) methods & technologies and gain an insight into the application areas of Natural language processing.

COURSE OUTCOMES:

After completing the course student should be able to:

1. Define different data models used in Information Retrieval using NLP.
2. Demonstrate current methods for statistical approaches to machine translation.
3. Apply syntactic parsing and semantic analysis on text.
4. Solve and implement real world problems using NLP.

Detailed Contents:

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,Viterbi algorithms and EM training.

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.Compositional semantics.

UNIT V:Application of NLP: intelligent work processors: Machine translation, user interfaces,Man-Machine interfaces, natural language querying, tutoring and authoring systems, speechrecognition, and commercial use of NLP.

Text Books:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication.
2. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, OReilly Media.
3. Manning and Schutze "Foundations of Statistical Natural Language Processing", MIT Press.

Reference Books:

1. Breck Baldwin, Language Processing with Java and LingPipe Cookbook, Atlantic Publisher.
2. Richard M Reese, Natural Language Processing with Java, OReilly Media.
3. Nitin Indurkhya and Fred J. Damerau, Handbook of Natural Language Processing, Chapman and Hall/CRC Press.
4. Tanveer Siddiqui, U.S. Tiwary, Natural Language Processing and Information Retrieval, Oxford University Press.