

Subject Code: 01CE0713**Subject Name: Natural Language****Processing B.Tech. Year - VIII**

Object Natural language processing deals with written text. Students will learn how to process written text from basic of fundamental knowledge starts with Finite automata, Regular expression and probabilistic model with n-grams. Recognizing Speech and parsing with grammar. This course also covers basis of semantic analysis and discourse analysis and drives it to machine translation. This NLP course will boost student knowledge to research level where they can conduct new level of research. It really helpful for undergraduate students.

Credits Earned: 4 Credits**Course Outcomes:** After completion of this course, student will be able to

- Learn and Understand Natural Language Processing (Comprehension).
- Probabilistic model of defining language and techniques.(Application)
- Applying Hidden Markov model and Speech Recognition.(Application)
- Application of context free grammar and language parsing.(Application)
- Implement probabilistic and language parsing.(Application)
- Differentiation of semantic and discourse in terms of NLP.(Analyse)

Pre-requisite of course: Data Structure, Theory of Computation, Compiler Design.

Teaching and Examination Scheme

Teaching Scheme (Hours)			Credits	Theory Marks			Tutorial/ Practical Marks		Total Marks
Theory	Tutorial	Practical		ESE (E)	Mid Sem (M)	Internal (I)	Viva (V)	Term work (TW)	
3	0	2	4	50	30	20	25	25	150

Contents:

Unit	Topics	Contact Hours
1	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	6
2	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.	8
3	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.	8
4	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.	12
5	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, BrownCorpus, British National Corpus (BNC).	8
	Total Hours	42

References:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
2. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, First Edition, O'Reilly Media, 2009.
3. Breck Baldwin, —Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015.
4. Richard M Reese, —Natural Language Processing with Java, O'Reilly Media,

5. Nitin Indurkha and Fred J. Damerau, —Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
6. Tanveer Siddiqui, U.S. Tiwary, “Natural Language Processing and Information Retrieval”, Oxford University Press, 2008.

Suggested Theory distribution:

The suggested theory distribution as per Bloom's taxonomy is as per follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process.

Distribution of Theory for course delivery and evaluation					
Remember	Understand	Apply	Analyse	Evaluate	Create
10%	20%	35%	15%	10%	10%

Instructional Method:

- a. The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.
- b. The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.
- c. Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.
- d. Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory

Supplementary Resources:

1. <https://blog.algorithmia.com/introduction-natural-language-processing-nlp/>
2. <https://www.udacity.com/course/natural-language-processing-nanodegree--nd892>
3. <https://www.coursera.org/learn/language-processing>
4. <https://towardsdatascience.com/a-practitioners-guide-to-natural-language-processing-part-i-processing-understanding-text-9f4abfd13e72>
5. <https://www.edx.org/course/natural-language-processing>

Practical Resources:

Any 10 practical related to text parsing, tokenization and semantic word fetching related. which is up to faculties. faculties can plan further practical based upon his/her knowledge.