

COURSE TITLE	NATURAL LANGUAGE PROCESSING
COURSE CODE	01AI0702
COURSE CREDITS	4

Objective:

- 1 Student will understand Computer Vision and their principles. The course will cover theory as well as practice aspects of a subject through scheduled lectures and labs, course will cover details of fundamentals of machine vision, features detection, object recognition and detection, segmentation and various applications.
- 2 To tag a given text with basic Language processing features, design an innovative application using NLP components, implement a rule based system to tackle morphology/syntax of a Language, design a tag set to be used for statistical processing keeping an application in mind, design a Statistical technique for a new application, Compare and contrast use of different statistical approaches for different types of applications.

Course Outcomes: After completion of this course, student will be able to:

- 1 To tag a given text with basic Language features
- 2 To design an innovative application using NLP components
- 3 To implement a rule based system to tackle morphology/syntax of a language
- 4 To design a tag set to be used for statistical processing for real-time applications
- 5 To compare and contrast the use of different statistical approaches for different types of NLP applications.

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

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	2	50	30	20	25	25

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	8

Contents : Unit	Topics	Contact Hours
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	10
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	10
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, Distributional methods	14
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, Brown Corpus, British National Corpus (BNC).	10
Total Hours		52

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	Experiments Word Analysis, Word Generation, Morphology, N-Grams, N-Grams Smoothing, POS Tagging: Hidden Markov Model, POS Tagging: Viterbi Decoding, Building POS Tagger, Chunking, Building Chunker	20
Total Hours		20

Textbook :

- 1 Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Daniel Jurafsky, James H. Martin, Pearson Publication,, 2014

References:

- 1 Daniel Jurafsky, James H. Martin?Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014

References:

- 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, 2015.
- 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 follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking / Creative
0.00	20.00	40.00	20.00	0.00	20.00

Instructional Method:

- 1 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
- 2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room
- 3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory
- 4 Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory
- 5 <https://nptel.ac.in/courses/106/105/106105158/>, A course on Natural Language Processing, by Prof. Pawan Agrawal, IIT Kharagpur

Supplementary Resources:

- 1 <https://nptel.ac.in/courses/106/105/106105158/>, A course on Natural Language Processing, by Prof. Pawan Agrawal, IIT Kharagpur
- 2 <https://nptel.ac.in/courses/106/106/106106211/>, A course on Applied Natural Language Processing, Prof. Rameshan Ramchandran, IIT Madras
- 3 NLP projects, <https://www.dezyre.com/projects/data-science-projects/nlp-projects/>