

**Year: B. Tech IV (Semester VII)**

**Subject Name:** Natural Language Processing  
**Type of course:** Professional Elective Course  
**Prerequisite (if any):** -

**Subject Code:** BTIT14706

**Rationale:** Natural language processing (NLP) is a technology that gives computers the ability to process, comprehend and generate human language. It has many useful applications like NLP enabled search engines, document classification, machine translation etc. The purpose of this course is to give students an understanding of important concepts in NLP and equip them with knowledge to perform various NLP tasks which they can apply to diverse applications.

**Teaching and Examination Scheme:**

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	0	3	60	25	15	0	0	100

CA1: Continuous Assessment (assignments / projects / open book tests / closed book tests) CA2: Sincerity in attending classes / class tests / timely submissions of assignments / self-learning attitude / solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work / Quality of work submitted / Active participation in lab sessions / viva on practical skills learned in course.

**Contents:**

Sr. No.	Contents	Total Hours
1.	<b>Introduction</b> Importance of Natural Language Processing, Language Models, Stages in Natural Language Processing: Tokenization, Lexical analysis, Syntactic Analysis, Semantic Analysis, Pragmatic analysis, Overview of applications, Issues in NLP- ambiguity in language, synonyms, homonyms	08
2.	<b>Text Pre-processing</b> Challenges to text pre-processing, Character-Set Dependence , Language Dependence, Corpus Dependence, Application Dependence; Tokenization, Sentence Segmentation	06
3.	<b>Lexical and Syntactic Analysis</b> Lexical Analysis- Finite State Morphology: Disjunctive Affixes, Inflectional Classes and Exceptionality, Difficult morphological analysis: Isomorphism problems, Contiguity problems; Paradigm based lexical analysis	07

	Syntactic Parsing- Introduction, LR Parsing, Parts-of-Speech tagging, Constraint based Grammars, Issues in Parsing	
4.	<b>Semantic Analysis</b> Basic concepts and Issues in Natural Language Semantics, Approaches to Semantic Representation, Relational Issues in Lexical Semantics	05
5.	<b>Empirical and Statistical Approaches</b> Corpus Creation, Treebank Annotation, Statistical Parsing: Generative and Discriminative Models, Hidden Markov Model and Viterbi Algorithm for Sequence Prediction, Multiword expressions, Normalized web distance and word similarity, word sense disambiguation	07
6.	<b>Pragmatics</b> Discourse, Dialogue and Conversational Agents, Natural Language Generation, Dictionary Based Approaches, Reference Resolution, Pronoun Resolution	05
7.	<b>Applications</b> Text classification, Spell-checking, Question-Answering, Text Generation, Chabot's, Machine Translation, Spam Filtering, NLP based contemporary Search Engines	07

**Suggested Specification table with Marks (Theory): (For B. Tech only)**

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
20	20	10	10	-	-

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (Revised Bloom's Taxonomy)

**Reference Books:**

Sr no	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication	Publication Edition
1	Handbook of Natural Language Processing	Nitin Indurkha, Fred J. Damerau	Chapman & Hall/ CRC Press	2010	2 <sup>nd</sup> Edition
2	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Pearson Education	2008	2 <sup>nd</sup> Edition
3	Natural Language Understanding	James Allen	Addison-Wesley	1994	2 <sup>nd</sup> Edition

**Course Outcomes (CO):**

Sr. No.	CO statements	Marks % weightage
CO-1	Explain the basic concepts of NLP and its issues	25%
CO-2	Apply acquired knowledge of NLP to text pre-processing and statistical analysis tasks	25%
CO-3	Analyze the syntax, semantics and pragmatics of a statement written in a natural language	25%
CO-4	Design solutions for NLP applications like text classification, question-answering and machine translation	25%

**List of Open learning website**

- NPTEL course on Natural Language Processing ([https://onlinecourses.nptel.ac.in/noc23\\_cs45](https://onlinecourses.nptel.ac.in/noc23_cs45))