COURSE CONTENT

NATURAL LANGUAGE PROCESSING								
II Semester: CSE								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
BCSE21	**Elective**	3	-	-	3	40	60	100
Contact Classes:48	**Total Tutorials: Nil**	**Total Practical Classes: Nil**				**Total Classes: 45**		
Prerequisite: Data structures and Algorithms, Machine learning								

I. COURSE OVERVIEW:

This course provides students with a fundamental understanding of the key concepts and techniques in natural language processing, including syntax, semantics, morphology, and pragmatics, methods for representing and preprocessing text data, including techniques like tokenization, stemming, lemmatization, and vectorization. These concepts equip students with the knowledge and skills needed to understand, implement, and contribute to the field of natural language processing.

II. COURSE OBJECTIVES:

The students will try to learn:

I. The problems and solutions of NLP and their relation to linguistics and statistics
II. The techniques for extracting semantic information from text and understanding discourse structures.
III. The discourse processing and structure modeling

III. COURSE OUTCOMES:

After successful completion of the course, students will be able to:

CO 1 Understand the morphological models and find the structure of documents
CO 2 Apply parsing techniques for syntax analysis and represents syntactic structure
CO 3 Summarize the semantic parsing in natural language parsing and understand system paradigm
CO 4 Analyze NLP algorithms Able to design different language modeling Techniques.
CO 5 Analyze the structure language modeling and understand the dis course processing.

IV.COURSE CONTENT:

MODULE-I: FINDING THE STRUCTURE OF WORDS (9)
Words and Their Components, Issues and Challenges, Morphological Models. Finding the Structure of Documents: Introduction, Methods, Complexity of the Approaches, Performances of the Approaches.

MODULE-II: SYNTAX ANALYSIS (9)
Parsing Natural Language, Treebanks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure, Parsing Algorithms, Models for Ambiguity Resolution in Parsing, Multilingual Issues.

MODULE-III: SEMANTIC PARSING (9)
Introduction, Semantic Interpretation, System Paradigms

Word Sense Systems, Software.

MODULE-IV: PREDICTATE ARGUMENT STRUCTURE (9)
Predicate-argument structure, Meaning Representation Systems, Software.

MODULE-V: DISCOURSE PROCESSING (9)
Cohesion, Reference Resolution, Discourse Cohesion and Structure Language Modeling: Introduction, N-Gram Models, Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems, Multilingual and Cross-Lingual Language Modeling.

V. TEXTBOOKS:

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd edition, 2014.

VI. REFERENCE BOOKS:

1. Speech and Natural Language Processing - Daniel Jura sky & James H Martin, Pearson Publications.

VII. WEB REFERENCES:

1. https://link.springer.com/book/10.1007/978-981-99-1999-4
2. https://freecomputerbooks.com/Natural-Language-Processing-Succinctly.html

VIII E-Text Books:

1. https://www.ai-startups.org/books/natural_language_processing/
2. https://tjzhifei.github.io/resources/NLTK.pdf
3. http://languagetechnologies.uohyd.ac.in/knm-publications/nlp-book.pdf