NATURAL LANGUAGE PROCESSING
[As per Choice Based Credit System (CBCS) scheme]
(Effective from the academic year 2017 - 2018)
SEMESTER – VII

Subject Code	17CS741	IA Marks	40
Number of Lecture Hours/Week	3	Exam Marks	60
Total Number of Lecture Hours	40	Exam Hours	03

CREDITS – 03

Module – 1	**Teaching Hours**
Overview and language modeling: Overview: Origins and challenges of NLP-Language and Grammar-Processing Indian Languages- NLP Applications-Information Retrieval. Language Modeling: Various Grammar- based Language Models-Statistical Language Model.	**8 Hours**
Module – 2	
Word level and syntactic analysis: Word Level Analysis: Regular Expressions-Finite-State Automata-Morphological Parsing-Spelling Error Detection and correction-Words and Word classes-Part-of Speech Tagging. Syntactic Analysis: Context-free Grammar-Constituency- Parsing-Probabilistic Parsing.	**8 Hours**
Module – 3	
Extracting Relations from Text: From Word Sequences to Dependency Paths: Introduction, Subsequence Kernels for Relation Extraction, A Dependency-Path Kernel for Relation Extraction and Experimental Evaluation. **Mining Diagnostic Text Reports by Learning to Annotate Knowledge Roles:** Introduction, Domain Knowledge and Knowledge Roles, Frame Semantics and Semantic Role Labeling, Learning to Annotate Cases with Knowledge Roles and Evaluations. **A Case Study in Natural Language Based Web Search:** InFact System Overview, The GlobalSecurity.org Experience.	**8 Hours**
Module – 4	
Evaluating Self-Explanations in iSTART: Word Matching, Latent Semantic Analysis, and Topic Models: Introduction, iSTART: Feedback Systems, iSTART: Evaluation of Feedback Systems, **Textual Signatures: Identifying Text-Types Using Latent Semantic Analysis to Measure the Cohesion of Text Structures:** Introduction, Cohesion, Coh-Metrix, Approaches to Analyzing Texts, Latent Semantic Analysis, Predictions, Results of Experiments. **Automatic Document Separation: A Combination of Probabilistic Classification and Finite-State Sequence Modeling:** Introduction, Related Work, Data Preparation, Document Separation as a Sequence Mapping Problem, Results. **Evolving Explanatory Novel Patterns for Semantically-Based Text Mining:** Related Work, A Semantically Guided Model for Effective Text Mining.	**8 Hours**
Module – 5	
INFORMATION RETRIEVAL AND LEXICAL RESOURCES: Information Retrieval: Design features of Information Retrieval Systems-Classical, Non classical, Alternative Models of Information Retrieval – valuation Lexical	**8 Hours**

Resources: World Net-Frame Net- Stemmers-POS Tagger- Research Corpora.	
Course outcomes: The students should be able to:	
• Analyze the natural language text. • Define the importance of natural language. • Understand the concepts Text mining. • Illustrate information retrieval techniques.	
Question paper pattern: The question paper will have ten questions. There will be 2 questions from each module. Each question will have questions covering all the topics under a module. The students will have to answer 5 full questions, selecting one full question from each module.	
Text Books:	
1. Tanveer Siddiqui, U.S. Tiwary, "Natural Language Processing and Information Retrieval", Oxford University Press, 2008. 2. Anne Kao and Stephen R. Poteet (Eds), "Natural LanguageProcessing and Text Mining", Springer-Verlag London Limited 2007.	
Reference Books:	
1. Daniel Jurafsky and James H Martin, "Speech and Language Processing: Anintroduction to Natural Language Processing, Computational Linguistics and SpeechRecognition", 2nd Edition, Prentice Hall, 2008. 2. James Allen, "Natural Language Understanding", 2nd edition, Benjamin/Cummingspublishing company, 1995. 3. Gerald J. Kowalski and Mark.T. Maybury, "Information Storage and Retrieval systems", Kluwer academic Publishers, 2000.	