

**NATURAL LANGUAGE PROCESSING**  
**[As per Choice Based Credit System (CBCS) scheme]**  
**(Effective from the academic year 2016 -2017)**

**SEMESTER – VII**

Subject Code	15CS741	IA Marks	20
Number of Lecture Hours/Week	3	Exam Marks	80
Total Number of Lecture Hours	40	Exam Hours	03

**CREDITS – 03**

**Course objectives:** This course will enable students to

- Learn the techniques in natural language processing.
- Be familiar with the natural language generation.
- Be exposed to Text Mining.
- Understand the information retrieval techniques

**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:**

**8 Hours**

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.

**Module – 4**

**Evaluating Self-Explanations in iSTART: Word Matching, Latent Semantic Analysis, and Topic Models:** Introduction, iSTART: Feedback Systems, iSTART: Evaluation of Feedback Systems,

**8 Hours**

**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.



<b>Module – 5</b>	
<b>INFORMATION RETRIEVAL AND LEXICAL RESOURCES:</b> Information Retrieval: Design features of Information Retrieval Systems-Classical, Non classical, Alternative Models of Information Retrieval – valuation Lexical Resources: World Net-Frame Net- Stemmers-POS Tagger- Research Corpora.	<b>8 Hours</b>
<b>Course outcomes:</b> The students should be able to:	
<ul style="list-style-type: none"> <li>Analyze the natural language text.</li> <li>Generate the natural language.</li> <li>Do Text mining.</li> <li>Apply information retrieval techniques.</li> </ul>	
<b>Question paper pattern:</b> 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.	
<b>Text Books:</b>	
<ol style="list-style-type: none"> <li>1. Tanveer Siddiqui, U.S. Tiwary, "Natural Language Processing and Information Retrieval", Oxford University Press, 2008.</li> <li>2. Anne Kao and Stephen R. Poteet (Eds), "Natural Language Processing and Text Mining", Springer-Verlag London Limited 2007.</li> </ol>	
<b>Reference Books:</b>	
<ol style="list-style-type: none"> <li>1. Daniel Jurafsky and James H Martin, "Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition", 2nd Edition, Prentice Hall, 2008.</li> <li>2. James Allen, "Natural Language Understanding", 2nd edition, Benjamin/Cummings publishing company, 1995.</li> <li>3. Gerald J. Kowalski and Mark.T. Maybury, "Information Storage and Retrieval systems", Kluwer academic Publishers, 2000.</li> </ol>	

  
H. O. D.

Dept. Of Computer Science & Engineering  
Alva's Institute of Engg. & Technology  
Mijar, MOODBIDRI - 574 226