||Jai Sri Gurudev||
Sri AdichunchanagiriShikshana Trust (R)

SJB Institute of Technology

Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi
Accredited by NAAC, New Delhi with 'A' Grade, Recognized by UGC, New Delhi with 2(f) and 12(B)
Certified by ISO 9001- 2015
BGS Health & Education City, Dr. Vishnuvardhan Road, Kengeri, Bengaluru – 560 060
Website:www.sjbit.edu.in, Email :principal@sjbit.edu.in, Mob:+**91-6366041109**

Department of Information Science and Engineering

Course Title: NATURAL LANGUAGE PROCESSING			**Course Code:** 18CS743
Semester: VII	**Academic Year:**21 – 22	**Total hrs.:**40	**Hrs./Week:** 03
Int. Exam Hrs.: 01	**Internal Evaluation Max. Marks:** 40		
Ext. Exam Hrs.: 03	**Ext. Exam Max.Marks:**60		
Lesson Plan Author / Desgn. / Dept.: CHETAN R/ Asst. Professor / IS&E			

Course objectives:
This course will enable students to:
- Define Natural Language Processing
- Explain Word level and syntactic analysis
- Analyze the natural language text
- Understand the concepts of Text mining
- Illustrate information retrieval techniques.

Course outcomes:
The students should be able to:

CO1	Apply the natural language text at word level and syntactic structures to develop simple and complex applications
CO2	Analyze the natural language text at word level and syntactic structures.
CO3	Identify the concepts of Text mining.
CO4	Apply information retrieval techniques to develop applications.

CO- PO Mapping
Course Name   - NATURAL LANGUAGE PROCESSING [18CS743]
Semester: VII   **Year of Study: 2021-22**

Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2	2											2	
CO2	2	2	2										2	
CO3	2	2	2										2	

CO4	3	3											2	
	2.25	2.25	2										2	

Syllabus

Course Title: Natural Language Processing   **Course Code:** 18CS743

MODULE-I	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**
Blooms Taxonomy:L1, L2, L3	
MODULE-II	Teaching Hours
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**
Blooms Taxonomy: L1, L2, L3	
MODULE-III	Teaching Hours
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**
Blooms Taxonomy: L1, L2, L3	
MODULE-IV	Teaching Hours
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**
Blooms Taxonomy: L1, L2, L3	
MODULE-V	Teaching Hours
INFORMATION RETRIEVAL AND LEXICAL RESOURCES: Information Retrieval: Design features of Information Retrieval Systems-Classical, Non classical, Alternative Models of Information Retrieval – valuation Lexical Resources: World Net-	**8**

Frame Net- Stemmers-POS Tagger- Research Corpora.	
Blooms Taxonomy: L1, L2, L3	

Subject objectives:

Day	Module No. & Title	SUB TOPICS	CO
Day 1	**MODULE-1** **Overview and language modeling**	Overview: Origins and challenges of NLP	CO1
Day 2		Language and Grammar	CO1
Day 3		Processing Indian Languages	CO1
Day 4		NLP Applications	CO1
Day 5		Information Retrieval: Language Modeling	CO1
Day 6		Various Grammar	CO1
Day 7		Language Models	CO1
Day 8		Statistical Language Model	CO1
Day 9	**MODULE -2** **Word Level and Syntactic Analysis**	Word Level Analysis- Regular Expressions	CO1
Day 10		Finite State Automata	CO1
Day 11		Morphological Parsing	CO2
Day 12		Spelling Error Detection and Correction	CO2
Day 13		Words and Word classes	CO2
Day 14		Part of Speech Tagging	CO2
Day 15		Syntactic Analysis: Context Free Grammar Consistency	CO2
Day 16		Parsing-Probabilistic Parsing	CO2
Day 17	**MODULE – 3** **Extracting Relations from Text, Mining Diagnostic Text Reports by Learning to Annotate Knowledge Roles,** **A Case Study in Natural Language Based Web Search** **Students will be able to learn:** **Students will know:**	Introduction, Subsequence Kernels for Relation Extraction	CO1
Day 18		A Dependency Path Kernel for Relation Extraction	CO1
Day 19		Experimental Evaluation	CO1
Day 20		Introduction, Domain Knowledge and Knowledge Roles	CO1
Day 21		Frame Semantics and Semantic Role Labeling	CO2
Day 22		Learning to Annotate Cases with Knowledge Roles and Evaluations.	CO2
Day 23		InFact System Overview	CO2
Day 24		The Global Security.org Experience	CO2
Day 25	**MODULE - 4** **Evaluating Self-Explanations in iSTART,** **Textual Signatures,**	Introduction, iSTART	CO3
Day 26		Feedback Systems, iSTART, Evaluation of Feedback Systems.	CO3
Day 27		Introduction, Cohesion, Coh-Metrix	CO3

Day 28	**Automatic Document Separation ,**	Approaches to Analyzing Texts, Latent Semantic Analysis	CO3
Day 29	**Evolving Explanatory Novel Patterns for Semantically-Based Text Mining**	Predictions, Results of Experiments.	CO3
Day 30		Introduction, Related Work, Data Preparation,	CO3
Day 31		Document Separation as a Sequence Mapping Problem, Results	CO3
Day 32		Related Work, A Semantically Guided Model for Effective Text Mining	CO3
Day 33	**MODULE – 5 INFORMATION RETRIEVAL AND LEXICAL RESOURCES**	Introduction, Design Features of Information Retrieval Systems	CO3
Day 34		Classical IR	CO3
Day 35		Non-Classical IR	CO4
Day 36		Alternative Models for IR	CO4
Day 37		Evaluation of IR	CO4
Day 38		WordNet	CO4
Day 39		FrameNet, Stemmers	CO4
Day 40		POS Tagger, Research Corpora	CO4

Reference / Text Book Details

Sl. No.	Name of Book	Author Name	Publication
1	Natural Language Processing and Information Retrieval	Tanveer Siddiqui, U.S. Tiwary,	Oxford University Press, 2008
2	Natural Language Processing and Text Mining	Anne Kao and Stephen R. Poteet	Springer-Verlag London Limited 2007
3	Speech and Language Processing: Anintroduction to Natural Language Processing, Computational Linguistics and SpeechRecognition	Daniel Jurafsky and James H Martin	2nd Edition, Prentice Hall, 2008
4	Natural Language Understanding	James Allen	2nd edition, Benjamin/Cummingspublishing company, 1995
5	Gerald J. Kowalski and Mark.T. Maybury	Information Storage and Retrieval systems	Kluwer academic Publishers, 2000.

Faculty in Charge
[CHETAN R]

HOD
Professor and Head, Dept. of ISE