

OBJECTIVES:

The student should be made to: Learn the techniques in natural language processing. • Be familiar with the natural language generation. • Be exposed to machine translation. • Understand the information retrieval techniques. •

UNIT I OVERVIEW AND LANGUAGE MODELING

8

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.

UNIT II WORD LEVEL AND SYNTACTIC ANALYSIS

9

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.

UNIT III SEMANTIC ANALYSIS AND DISCOURSE PROCESSING

10

Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion-Reference Resolution- Discourse Coherence and Structure.

UNIT IV NATURAL LANGUAGE GENERATION AND MACHINE TRANSLATION

9

Natural Language Generation: Architecture of NLG Systems- Generation Tasks and Representations Application of NLG. Machine Translation: Problems in Machine Translation- Characteristics of Indian Languages- Machine Translation Approaches-Translation involving Indian Languages.

UNIT V INFORMATION RETRIEVAL AND LEXICAL RESOURCES

9

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.

TOTAL: 45 PERIODS

OUTCOMES: Upon completion of the course, the student should be able to: Analyze the natural language text. • Generate the natural language. • Do machine translation. • Apply information retrieval techniques. •

TEXT BOOK: 1. Tanveer Siddiqui, U.S. Tiwary, “Natural Language Processing and Information Retrieval”, Oxford University Press, 2008.

REFERENCES: 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. 2. James Allen, “Natural Language Understanding”, 2nd edition, Benjamin /Cummings publishing company, 1995. CS6012