MCAE316: NATURAL LANGUAGE PROCESSING

Hours/Week: 3	I.A. Marks: 30
Credits: 3	Exam. Marks: 70

Course Learning Objectives: Students will able to try,
1. To introduce students the challenges of empirical methods for natural language processing (NLP) applications.
2. To introduce basic mathematical models and methods used in NLP applications to formulate computational solutions.
3. To introduce students research and development work in information retrieval, information extraction, and knowledge discovery using different natural language resources.
4. Understand the principles of language resource annotation and its use in machine learning applications and apply the above principles in analysis of data and acquire intended information through the use of available tools.

_____

Course Outcomes: After completing the course, the students will be able to,
CO1: Understand basic approaches to syntax and semantics in NLP.
CO2: Realize approaches to discourse, generation and dialogue in NLP
CO3: Familiarize the current methods for statistical approaches to machine translation.
CO4: Understand machine learning techniques used in NLP, including hidden Markov models and probabilistic context-free grammar.
CO5: Familiar with clustering and unsupervised methods, log-linear and discriminative models, and the EM algorithm as applied within NLP
CO6: Understand the design and implementation issues in various NLP applications such as information retrieval and information extraction.
CO7: Understand the principles of language resource annotation and its use in machine learning applications and apply the above principles in analysis of data and acquire intended information through the use of available tools.

_____

UNIT-I   9 Hrs.
Overview and Language Modeling: 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   9 Hrs.
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.

UNIT-III   9 Hrs.
Semantic Analysis and Discourse Processing: Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion-Reference Resolution- Discourse Coherence and Structure.

UNIT-IV **9 Hrs.**

Natural Language Generation and Machine Translation: 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.

REFERENCE BOOKS:

1. Edward Loper, Ewan Klein, and Steven Bird, Natural Language Processing with Python, 'Reilly Publications, 2009.

2. Christopher D. Manning, Hinrich Schütze, Foundations of Statistical Natural Language Processing, MITpress, 1999.

3. Dan Jurafsky, James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall, 2009.