



MARRI LAXMAN REDDY INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(AN AUTONOMOUS INSTITUTION)

(Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad)

Accredited by NAAC with 'A' Grade & Recognized Under Section 2(f) & 12(B) of the UGC act, 1956

COURSE CONTENT

NATURAL LANGUAGE PROCESSING

VI Semester : CSD

VII Semester : CSE

VIII Semester : CSM

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P		CIA	SEE	Total
24X0530	Elective	3	0	0	3	40	60	100
		Contact Classes: 45		Tutorial Classes: Nil		Practical Classes: Nil		Total Classes: 45

Prerequisites: Data structures, compiler design

Course Overview:

The course examines the structural, syntactic, and semantic intricacies of natural language through advanced analytical and modeling paradigms. It further explores ambiguity resolution, semantic representation, and refined language modeling techniques such as n-gram and Bayesian frameworks.

Course Objectives:

1. To understand the structure of words and documents and explore morphological analysis techniques.
2. To study syntax and parsing of natural language using data-driven approaches and treebanks.
3. To address ambiguity resolution in parsing and analyze multilingual syntax issues.
4. To explore semantic parsing, predicate-argument structures, and meaning representation systems.
5. To learn language modeling techniques including n-gram models, Bayesian estimation, and multilingual adaptations.

Course Outcomes: After Completion of the Course, Students should be able to

1. Analyze the structure of words and documents using morphological models and feature extraction techniques.
2. Apply parsing algorithms and treebank approaches for syntactic analysis of natural language.
3. Resolve ambiguity in parsing and handle multilingual syntax challenges.
4. Develop semantic parsing models and represent meaning using predicate-argument structures.
5. Build language models using n-grams, Bayesian methods, and multilingual adaptations.

Module I

[10]

Finding the Structure of Words: Words and their components, issues and challenges, morphological models. Finding the Structure of Documents: Introduction, methods, complexity of approaches, performance of approaches, features.

Module II

[10]

Syntax I: Parsing natural language, treebanks as a data-driven approach to syntax, representation of syntactic structure, parsing algorithms.

Module III**[10]**

Syntax II: Models for ambiguity resolution in parsing, multilingual issues.

Semantic Parsing I: Introduction, semantic interpretation, system paradigms, word sense.

Module IV**[8]**

Semantic Parsing II: Predicate-argument structure, meaning representation systems.

Module V**[10]**

Language Modeling: Introduction, n-gram models, language model evaluation, Bayesian parameter estimation, language model adaptation, class-based language models, variable-length models, Bayesian topic-based models, multilingual and cross-lingual language modeling.

TEXT BOOKS:

1. Multilingual Natural Language Processing Applications: From Theory to Practice – Daniel M. Bikel and Imed Zitouni, Pearson Publications.

REFERENCE BOOKS:

1. Speech and Natural Language Processing – Daniel Jurafsky and James H. Martin, Pearson Publications.
2. Natural Language Processing and Information Retrieval – Tanvier Siddiqui and U.S. Tiwary.

ELECTRONIC RESOURCES:

1. <https://www.w3schools.com/search/?q=natural+language+processing>
2. <https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/>
3. <https://www.programiz.com/search?query=natural+language+processing>
4. https://www.tutorialspoint.com/natural_language_processing/index.htm
5. <https://www.codechef.com/practice>

MATERIALS ONLINE:

- Course template
- Tutorial question bank
- Tech talk and Concept Video topics
- Open-ended experiments
- Definitions and terminology
- Assignments
- Model question paper – I
- Model question paper – II
- Lecture notes
- E-Learning Readiness Videos