INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

TEXT ANALYTICS AND LANGUAGE PROCESSING								
VIII Semester: CSE								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACSE38	**Elective**	3	0	0	3	40	60	100
Contact Classes: 48	**Tutorial Classes: Nil**	**Practical Classes: Nil**				**Total Classes: 48**		
Prerequisite: Machine Learning								

I. COURSE OVERVIEW:

This course is study of computing systems that can process, understand, or communicate in human language. The primary focus of the course will be on understanding various NLP tasks and algorithms for effectively solving these problems, and methods for evaluating their performance. This course is intended as a theoretical and methodological introduction to a the most widely used and effective current techniques, strategies and toolkits for natural language processing, with a primary focus on those available in the Python programming language.

II. COURSES OBJECTIVES:

The students will try to learn

I. Natural language Processing for analyzing words based on Morphology and CORPUS.
II. Probability theory with Linguistic essentials such as syntactic and semantic analysis of text.
III. The applications of statistical learning methods and cutting-edge research models from deep learning.

III. COURSE OUTCOMES:

At the end of the course students should be able to

CO1 Remember the knowledge of complex language behavior in terms of phonetics, morphology
CO2 Understand the semantics and pragmatics for text processing.
CO3 Apply the CORPUS linguistics to compile and analyze the texts based on digestive games.
CO4 Understand various statistical approaches to machine translation for a given natural language.
CO5 Apply Part-of-speech (POS) tagging for a given natural language and suitable modeling technique based on the structure
CO6 The art algorithms and techniques for text-based processing of natural language with respect to morphology.

IV. COURSE CONTENT:

MODULE I: NATURAL LANGUAGE PROCESSING (09)

Introduction, Applications or Use cases of NLP, Components of NLP, Steps in NLP, Finding the Structure of Words: Words and Their Components, Lexemes, Morphemes, Morphology, Problems in morphological processing, Typology, Morphological Typology.

MODULE II: REGULAR EXPRESSIONS, TEXT NORMALIZATION (09)

Regular Expressions, patterns, words, Corpora, Text normalization, Minimum edit distance, Regular Language and FSAs, Raw Text Extraction and Tokenization, Extracting Terms from Tokens, Normalization.

MODULE III: N-GRAM LANGUAGE MODELS (09)

N-grams, Evaluating language models, Generalization and zeros, smoothing, kneser-Ney smoothing, huge language models and stupid back off. Perplexity's relation to entropy.

Inflection, Derivational Morphology, Finite-State Morphological Parsing, The Lexicon and Morphotactics, Morphological Parsing with Finite State Transducers, Combining FST Lexicon and rules.

MODULE IV: WORD SENSE DISAMBIGUATION (09)

Methodological Preliminaries, Supervised Disambiguation: Bayesian classification, An information theoretic approach, Dictionary-Based Disambiguation: Disambiguation based on sense, Thesaurus based disambiguation, Disambiguation based on translations in a second-language corpus.

MODULE V: MARKOV MODEL AND POS TAGGING (09)

Markov Model: Hidden Markov model, Fundamentals, Probability of properties, Parameter estimation,Variants, Multiple input observation. The Information Sources in Tagging: Markov model taggers, Viterbi algorithm, Applying HMMs to POS tagging, Applications of Tagging.

IV. TEXT BOOKS:

1. Christopher D. Manning and Hinrich Schutze, " Foundations of Natural Language Processing" , 6th Edition, The MIT Press Cambridge, Massachusetts London, England, 2003.
2. Daniel Jurafsky and James H. Martin "Speech and Language Processing", 3rd edition, Prentice Hall, 2009.

VII. REFERENCE BOOKS:

1. Nitin Indurkhya, Fred J. Damerau "Handbook of Natural Language Processing", Second Edition, CRC Press,2010.
2. James Allen "Natural Language Understanding", Pearson Publication 8th Edition. 2012.
3. Chris Manning and Hinrich Schütze, "Foundations of Statistical Natural Language Processing",2nd edition, IT Press Cambridge, MA, 2003.

VIII. ELECTRONICS RESOURCES:

1. https://www.academia.edu/7452675/Foundations_of_Statistical_Natural_Language_Processing
2. https://www.mygreatlearning.com/blog/natural-language-processing-tutorial/
3. https://pub.towardsai.net/natural-language-processing-nlp-with-python-tutorial-for-beginners-1f54e610a1a0
4. https://www.analyticsvidhya.com/blog/2021/02/basics-of-natural-language-processing-nlp-basics/
5. https://towardsdatascience.com/free-hands-on-tutorials-to-get-started-in-natural-language-processing-6a378e24dbfc

IX. MATERIALS ONLINE

1. Course Outline Description
2. Lecture notes
3. PowerPoint presentation
4. Definitions and Terminology
5. Tutorial Question Bank
6. Case Studies

7. Real life Examples
8. Complex Engineering Problems
9. Tech Talk Topics
10. Concept Video Topics
11. Open-ended Exercises
12. Assignments
13. Model Question Paper – I
14. Model Question Paper – II
15. GATE Question Bank
16. Previous Question Papers and Solutions