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

COURSE CONTENT

LANGUAGE MODELLING								
VII Semester: CSE (AI&ML)								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACAD40	Elective	3	0	0	3	40	60	100
Contact Classes: 48	**Tutorial Classes: Nil**	**Practical Classes: Nil**				**Total Classes: 48**		
Prerequisites: Python Programming / Applied Artificial Intelligence								

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 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. COURSE OBJECTIVES:

The students will try to learn:

I. The concepts and techniques of Natural language Processing for analyzing words based on Morphology and CORPUS.

II. The mathematical foundations, 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, student should be able to:

CO1 Remember the knowledge of complex language behavior in terms of phonetics, morphology etc,

CO2 Understand semantics and pragmatics for text processing.

CO3 Apply the CORPUS linguistics to compile and analyze the texts based on digestive approach (Text Corpus Method).

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 Apply state-of-the-art algorithms and techniques for text-based processing of natural language with respect to morphology.

IV. COURSE CONTENT:

MODULE – I: CLASSICAL APPROACH TO NLP (09)
Introduction to NLP, the classical tool kit, Knowledge in Speech and Language processing, ambiguity and models and algorithm, language and understanding, brief history.

MODULE – II: REGULAR EXPRESSIONS AND 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 (10)
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 Morphotactic, Morphological Parsing with Finite State Transducers, Combining FST Lexicon and rules.

MODULE – IV: WORDSENSE DISAMBIGUATION (10)
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 (10)
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.

V. TEXTBOOKS:

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.

VI. 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.

VII. WEB REFERENCES

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-processing6a378e24dbfc

VIII. MATERIALS ONLINE

1. Course template
2. Tutorial question bank

3. Tech talk topics
4. Open-ended experiments
5. Definitions and terminology
6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. PowerPoint presentation
11. E-Learning Readiness Videos (ELRV)