

Subject Code	Subject Name (Lab oriented Theory Courses)	Category	L	T	P	C
AI19643	FOUNDATIONS OF NATURAL LANGUAGE PROCESSING	PC	3	0	2	4

Objectives:						
●	To understand the concepts and techniques of Natural language Processing for analyzing words based on Morphology and CORPUS.					
●	To learn mathematical foundations, Probability theory with Linguistic essentials such as syntactic analysis of text.					
●	To learn mathematical foundations, Probability theory with Linguistic essentials such as and semantic analysis of text.					
●	To familiarize the Statistical learning methods from deep learning.					
●	To interpret cutting-edge research models from deep learning.					

<b>UNIT-I</b>	<b>INTRODUCTION TO NLP</b>	<b>9</b>
Introduction to NLP - Various stages of NLP –The Ambiguity of Language: Why NLP Is Difficult Parts of Speech: Nouns and Pronouns, Words: Determiners and adjectives, verbs, Phrase Structure. Statistics Essential Information Theory : Entropy, perplexity, The relation to language, Cross entropy( <b>T1:Chapters 1,2,3</b> )		
<b>UNIT-II</b>	<b>TEXT PREPROCESSING AND MORPHOLOGY</b>	<b>9</b>
Character Encoding, Word Segmentation, Sentence Segmentation, Introduction to Corpora, Corpora Analysis. Inflectional and Derivation Morphology, Morphological analysis and generation using Finite State Automata and Finite State transducer ( <b>T1:Chapters 4</b> )		
<b>UNIT-III</b>	<b>LANGUAGE MODELLING</b>	<b>9</b>
Words: Collocations- Frequency-Mean and Variance –Hypothesis testing: The t test, Hypothesis testing of differences, Pearson’s chi-square test, Likelihood ratios. Statistical Inference: n -gram Models over Sparse Data: Bins: Forming Equivalence Classes- N gram model - Statistical Estimators- Combining Estimators( <b>T1:Chapters 5, 6</b> )		
<b>UNIT-IV</b>	<b>WORD SENSE DISAMBIGUATION</b>	<b>9</b>
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.( <b>T1:Chapters 7</b> )		
<b>UNIT-V</b>	<b>MARKOV MODEL AND POS TAGGING</b>	<b>9</b>
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. ( <b>T1:Chapters 9,10</b> )		
<b>Contact Hours</b>		<b>: 45</b>

List of Experiments						
1.	Perform Morphological Analysis for an interrogative sentence, declarative sentence and complex sentences with more two sentences connected using conjunctions.					
2.	Perform Coarse-grained POS Tagging and Fine-grained POS Tagging.					
3.	Named Entity Recognition with Python.					
4.	Sentiment Analysis with Python.					
5.	Keyword Extraction with Python.					
6.	Spelling Correction Model with Python.					
7.	Resume Screening with Python.					
8.	Twitter Sentiment Analysis.					
9.	NLP For WhatsApp Chats.					
10.	NLP for Other languages.					
<b>Contact Hours</b>		<b>: 30</b>				
<b>Total Contact Hours</b>		<b>: 75</b>				

<b>Course Outcomes:</b>	
On completion of the course, the students will be able to	
●	Realize semantics and pragmatics of English language for text processing
●	Create CORPUS linguistics based on digestive approach (Text Corpus method)
●	Check a current methods for statistical approaches to machine translation.
●	Demonstrate the state-of-the-art algorithms and techniques for text-based processing of natural language with respect to morphology.
●	Develop a Statistical Methods for Real World Applications and explore deep learning based NLP.

<b>Text Books:</b>	
1	Christopher D. Manning and HinrichSchutze, “ Foundations of Natural Language Processing” , 6th Edition, The MIT Press Cambridge, Massachusetts London, England, 2003 2009.
2	Daniel Jurafsky and James H. Martin “Speech and Language Processing”, 3rd edition, Prentice Hall, 2009.

<b>Reference Books:</b>	
1	NitinIndurkha, 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 HinrichSchütze, “Foundations of Statistical Natural Language Processing”, 2nd edition, MITPress Cambridge, MA, 2003.
4	Hobson lane, Cole Howard, Hannes Hapke, “Natural language processing in action” MANNING Publications, 2019.
5	Alexander Clark, Chris Fox, Shalom Lappin, “The Handbook of Computational Linguistics and Natural Language Processing”, Wiley-Blackwell, 2012
6	Rajesh Arumugam, RajalingappaShanmugamani “Hands-on natural language processing with python: A practical guide to applying deep learning architectures to your NLP application”. PACKT publisher, 2018.

**Web link:**

1. <https://blog.algorithmia.com/introduction-natural-language-processingnlp>
2. <https://www.udacity.com/course/natural-language-processingnanodegree--nd892>
3. <https://www.coursera.org/learn/language-processing>
4. <https://towardsdatascience.com/a-practitioners-guide-to-naturallanguage-processing-part- processing-understanding-text-9f4abfd13e72>
5. <https://www.edx.org/course/natural-language-processing>