



Kadi Sarva Vishwavidyalaya
Faculty of Engineering & Technology
Fourth Year Bachelor of Engineering (Computer/IT)
 (To be Proposed For: Academic Year 2020-21)

Subject Code: CT703C-N	Subject Title: Natural Language Processing
Pre-requisite	Basics of Probability and Statistics, Linear Algebra, Data Structure and Algorithm

Teaching Scheme (Credits and Hours)

Teaching scheme				Total Credit	Evaluation Scheme					Total Marks
L	T	P	Total		Theory		Mid Sem Exam	CIA	Pract.	
Hrs	Hrs	Hrs	Hrs		Hrs	Marks	Marks	Marks	Marks	
04	00	02	06	05	3	70	30	20	30	150

Learning Objectives:

- Learn to analyse textual data
- Learn various methods to represent text in vector form
- Explore various application of NLP like Machine Translation, Text Summarization, Dialog system

Outline of the Course:

Sr. No	Title of the Unit	Minimum Hours
1	Introduction to NLP	6
2	N-Gram Language Model	6
3	Text Representation	10
4	Text classification and clustering	8
5	MORPHOLOGY AND PART OF SPEECH TAGGING	12
6	Text Parsing	7
7	Semantic Analysis	5
8	NLP Applications	10

Total hours (Theory): 64

Total hours (Lab): 32

Total hours: 96



Kadi Sarva Vishwavidyalaya
Faculty of Engineering & Technology
Fourth Year Bachelor of Engineering (Computer/IT)
(To be Proposed For: Academic Year 2020-21)

Detailed Syllabus:

Sr. No	Topic	Lecture Hours	Weight age(%)
1	Introduction to NLP <ul style="list-style-type: none">• Introduction to NLP• Text pre-processing :• Regular Expression, tokenization• Stemming• Minimum Edit distance	6	9
2	N-Gram Language Model <ul style="list-style-type: none">• Intro. N-Gram• N-Gram probability estimation and perplexity• Smoothing technique(Laplace/good Turing/Kneser-Ney/Interpolation	6	9
3	Text Representation <ul style="list-style-type: none">• Bag-of-word :Tf/IDF,Count vector• Vector space Model• Latent semantic Analysis• Word embedding• Word2Vec• Glove• fastText• Sentence embedding Technique: Doc2Vec	10	16
4	Text classification and clustering <ul style="list-style-type: none">• The text classification problem• Feature Selection• Naive Bayes text classification• k- nearest neighbors• Support vector Machine• Flat Clustering• K-means algorithm• Hierarchical clustering	8	12



KadiSarvaVishwavidyalaya
Faculty of Engineering & Technology
Fourth Year Bachelor of Engineering (Computer/IT)
 (To be Proposed For: Academic Year 2020-21)

5	<p>MORPHOLOGY AND PART OF SPEECH TAGGING</p> <ul style="list-style-type: none"> • Morphology • Part of speech Tagging - Rule-Based Part of Speech Tagging • Markov Models - Hidden Markov Models • Viterbi algorithms • Maximum Entropy Models. <p>Sequence Processing with Recurrent Networks</p> <ul style="list-style-type: none"> • Simple Recurrent Neural Networks • Applications of Recurrent Neural Networks • Deep Networks: Stacked and Bidirectional RNNs • Managing Context in RNNs: LSTMs and GRUs • Words, Subwords and Characters 	12	19
6	<p>Text Parsing</p> <ul style="list-style-type: none"> • Syntax Parsing • Grammar formalisms and treebanks . • Parsing with Context Free Grammars • Features and Unification • Statistical parsing and probabilistic CFGs (PCFGs) 	7	11
7	<p>Semantic Analysis</p> <ul style="list-style-type: none"> • Lexical semantics and word-sense disambiguation. • Compositional semantics. 	5	8
8	<p>NLP Applications</p> <ul style="list-style-type: none"> • Information Extraction, Introduction to Named Entity Recognition and Relation Extraction • Question Answering • Text Summarization • Dialog System : • Machine Translation 	10	16

Instructional Method and Pedagogy:

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lecture and laboratory which carries 10 marks in overall evaluation.
- One internal exam will be conducted as a part of internal theory evaluation.
- Assignments based on the course content will be given to the students for each unit and will be evaluated at regular interval evaluation.
- Surprise tests/Quizzes/Seminar/tutorial will be conducted having a share of five marks in the overall internal evaluation.



Kadi Sarva Vishwavidyalaya
Faculty of Engineering & Technology
Fourth Year Bachelor of Engineering (Computer/IT)
(To be Proposed For: Academic Year 2020-21)

- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Experiments shall be performed in the laboratory related to course contents.

Learning Outcome:

- To Understand Document as Vector
- Various Supervised and Unsupervised learning Method
- Basic technique for language processing
- Text analysis
- Machine translation

Text Book:

- D. Jurafsky and J. Martin “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition”, Third Edition draft
- C. Manning and H. Schutze, “Foundations of Statistical Natural Language Processing”, MIT Press, 1999

List of Practical:

Sr. No	Name of Experiment
1	Basic Text Processing operation on text document.
2	Implement N-gram Language model
3	Write a program to extract features from text
4	Implement word embedding using Word2Vec/Glove/fastText
5	Implement LSA and Topic model.
6	Implementation text classification using Naïve Bayes, SVM.
7	Implementation of K-means Clustering algorithm on text.
8	Implement PoS Tagging on text
9	Implement text processing with neural network
10	Implement text processing with LSTM
11	Implement HMM/CRF on sequence tagging task
12	Develop any one NLP application Sentiment Analysis <ul style="list-style-type: none">• Chatbot• Text Summarization Track• Machine Translation• Question/Answering