SANKALCHAND PATEL UNIVERSITY
।। अथातो ज्ञानजिज्ञासा ।।

Ph. D. Program
(July-2019 Batch)

Subject Code: 1RI3000102	Subject Title: Essentials of Disciplinary Research Subject-1 Subject Sub-Title: Natural Language Processing
Pre-requisite Subject	

Objectives of course:
The objective of the course is to understand the concepts of morphology, syntax, semantics and pragmatics of natural language. This course covers computational properties of natural language and the commonly used algorithms for processing linguistic information. It also examines NLP models and algorithms for both traditional and statistical approach.

Learning outcomes:
At the completion of this course, the student will be able to:

- Understand the approaches to syntax and semantics in NLP
- Understand the approaches to discourse generation, summarization in NLP
- Understand the statistical approaches in NLP
- Understand Machine Learning techniques used in NLP, including Hidden Markov Model and probabilistic context free grammar, clustering and unsupervised methods, discriminative models.

Teaching Scheme (Credit Distribution)				Assessment Scheme (Weightage)	
Lecture	Tutorial	Practical	Credit	Hall Exam	Continuous Evaluation
4	--	--	4	60%	40%

Continuous evaluation is based on assignments/quiz/presentation/small projects as per the supervisor instructions.

	Subject Content		
Unit	Topic	No. of Lecture	Weightage %
1	**Introduction and Basic Text Processing** Introduction to NLP, Why is NLP hard, Empirical Laws, Text Processing Basics	2	10%
2	**Spelling Correction** Spelling Correction- Edit Distance, Weighted Edit Distance and other variations, Noisy Channel Model for spelling correction, N-gram Language Models, Evaluation of Language Models and basic smoothing, Language Modeling Advance : Advance Smoothing, Computational Morphology, Finite State Methods for Morphology	4	20%
3	**Part-of-Speech Tagging** Introduction to POS Tagging, Hidden Markov Models for POS Tagging, Viterbi Decoding for HMM, Parameter Learning, Baum Welch Algorithm, Maximum Entropy Models, Conditional Random Fields	3	15%

4	**Syntax** Syntax Parsing, CKY, PCFGs, Inside-Outside Probabilities, Dependency Grammars and Parsing, Transition-based parsing: Formulation, Transition-based parsing: Learning, MST-based Dependency Parsing	3	15%
5	**Topic Models** Topic Models: Introduction, Latent Dirichlet Allocation: Formulation, Gibbs Sampling for LDA, LDA Variants, Entity Linking, Information Extraction Introduction, Relation Extraction, Distant Supervision	3	15%
6	**Text Summarization, Text Classification using Python Programming** Stages of Text Summarization - Content Selection, Information Ordering, Sentence Realization, Removing Redundancy Text Summarization and Text Classification Using Python libraries – Pandas, Scikit-learn, Keras, TextBlob- Dataset Preparation, Feature Engineering, Model Training	4	15%
7	**Sentiment Analysis and Opinion Mining using Python Programming** Introduction, Applications of Sentiment Analysis, Challenges, Features of Sentiment Analysis, Machine Learning Approaches, Cognitive Approaches . Using Python's Natural Language Toolkit to solve problems in Sentiment Analysis	2	10%

List of References:

1. Dan Jurafsky and James Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, Second Edition, 2009.

2. Chris Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA: May 1999.

SANKALCHAND PATEL UNIVERSITY
।। अथातो ज्ञानजिज्ञासा ।।

Ph. D. Program
(July-2019 Batch)

Subject Code: 1RI3000103	**Subject Title: Essentials of Disciplinary Research Subject-2** **Subject Sub-Title: Natural Language Processing Python libraries**
Pre-requisite Subject	

Objectives of course:
The objective of the course is to understand various concepts of natural language processing along with their implementation and learn different modules available in Python for NLP. This course is designed with the aim to develop the expertise in building NLP applications.

Learning outcomes:
At the completion of this course, the students will able to:

- Implement the reading and working with text data using Python
- Implement the Text Pre-processing task of POS Tagging, NER using the NLTK and spaCy libraries
- Implement Stemming and Lemmatization using Python Libraries
- Implement topic modeling of the like LDA, LSA using the Python libraries
- Implement the text classification algorithms
- Implement deep learning algorithms for NLP

Teaching Scheme (Credit Distribution)				**Assessment Scheme (Weightage)**	
Lecture	**Tutorial**	**Practical**	**Credit**	**Hall Exam**	**Continuous Evaluation**
2	--	--	2	60%	40%

Continuous evaluation is based on assignments/quiz/presentation/small projects as per the supervisor instructions.

Subject Contents			
Unit	**Topic**	**No. of Lecture**	**Weightage %**
1	**The study of NLP Python Libraries** spaCy, Gensim, Pattern, NLTK, TextBlob, Polyglot, Vocbulary, PyNLPI, Stanford CoreNLP Python, MontyLingua	2	10%
2	**NLP using Python and the NLTK Toolkit** Installing NLTK Toolkit package in Python, Importing the dataset, Tokenization, Removing stop words, lemmatization and stemming, POS Tagging, Vocabulary of a dictionary	3	20%
3	**NLP using TextBlob** Tokenization, Noun Phrase extraction, POS Tagging, Words Inflection and Lemmatization, N-grams, Sentiment Analysis, Spelling Correction, Creating a short summary of text, Translation and Language detection, Text Classification using TextBlob	3	20%

4	**NLP using spaCy** Tokenization, POS Tagging, Entity Detection, Dependency Parsing, Noun Phrases, Word Vectors, Integrating spaCy with Machine Learning	3	15%
5	**Stanford CoreNLP Python** Setting up StanfordNLP in Python, Tokenization, Multi-word Token Expansion, Lemmatization, POS Tagging, Dependency Parsing	2	15%
6	**Deep Learning for NLP** Pre-processing natural language for use in machine learning applications, transform natural language into numerical representations with word2vec, make predictions with deep learning models trained on natural language, apply advanced NLP approaches with Keras, the high-level TensorFlow API, improve deep learning model performance by tuning hyperparameters	4	20%

List of References:

1. Steven Bird, Ewan Klein, Edward Loper: Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit ,Oreilly, June 2009.
2. Dan Jurafsky and James Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, Second Edition, 2009.
3. Chris Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA: May 1999.