

<b>Program Name</b>	<b>B. Tech. (Computer Engineering)</b>	<b>Semester – VII</b>
<b>Course Code</b>	<b>R4CO4204P</b>	
<b>Course Title</b>	<b>Natural Language Processing Lab</b>	
<b>Prerequisite</b>	<b>Machine Learning</b>	

#### **COURSE OUTCOMES: Students will be able to**

1.	Perform and understand stages of NLP and their application using NLTK toolkit
2.	Use and processing of corpora, word embeddings in NLP experiments.
3.	Access and analyse data from social platforms and perform text classification.
4.	Implement Different approaches such as Supervised, Unsupervised and Knowledge-based approaches to NLP applications

#### **COURSE CONTENTS**

	<b>Hrs</b>	<b>CO</b>
1. Use of Natural Language Toolkit (NLTK): Computing with Language: Texts and Words, Tokenization, Segmentation, Texts as Lists of Words, Simple Statistic Generation	2	1
2. Accept the Sentence as input and display Part of Speech Tags for the same. Also Perform dependency parsing and list out all dependency relations along with their operands/entities.	2	1
3. A word can be simple or complex. For example, the word 'cat' is simple because one cannot further decompose the word into smaller parts. On the other hand, the word 'cats' is complex, because the word is made up of two parts: root 'cat' and plural suffix '-s'. Develop an Analyser/Generator with Simple User Interface to accept lanua as English / Hindi and a word as input. Now generate the following information for the same. Word, Root, PoS tag, Gender and Person (if noun), Case and Tense (if Verb)	2	1
4. Probability of a sentence can be calculated by the probability of sequence of words occurring in it. We can use Markov assumption, that the probability of a word in a sentence depends on the probability of the word occurring just before it. Such a model is called a first order Markov model or the bigram model. Create a Bigram Language Model using given Corpora and generate the best possible word after a word (given as input).	2	1, 2
5. Text Processing at the Lowest Level, Text Processing with Unicode, Regular Expressions for Detecting Word Patterns, Useful Applications of Regular Expressions, Normalizing Text, Regular Expressions	2	1
6. Develop a Knowledge-based approach for Disambiguating the meaning of an input word from a given sentence.	2	2, 4
7. Experiment to understand and illustrate the word-vectors	2	1, 2
8. Develop a program to extract tweets from twitter platforms, Preprocess the data and store back the cleaned data in a .csv which related to a particular domain or specific hashtag value file.	2	3
9 Case study on Text Classification using supervised models	4	3,4
10 Case study on Text classification using Neural Network approach such as LSTM networks.	4	3, 4

## **TEXTBOOKS**

1. Steven Bird, Ewan Klein and Edward Loper, “Natural Language Processing with Python”, O'Reilly Media, 2009.
2. Joseph D. Booth, “Natural Language Processing, Syncfusion”, Inc., 2018

## **RECOMMENDED READING**

- 1 Dan Jurafsky, James H. Martin, “Speech and Language Processing, Stanford University”, 2017.
- 2 Shuly Wintner, “Formal Language Theory for Natural Language Processing”, ESSLLI, 2001.
- 3 Nitin Indurkha and Fred J. Damerau, “Handbook of Natural Language Processing”, Second Edition, Chapman and Hall/CRC Press, 2010.
- 4 Tanveer Siddiqui, U.S. Tiwary, “Natural Language Processing and Information Retrieval”, Oxford University Press, 2008.