

| Course Title | Natural Language Processing in AI (PE-IV) | | | | Course Type | | Theory | |
|------------------|-------------------------------------------|----------|---------------|-----------|-----------------|-----------|-------------------------|------------|
| Course Code | B22EAS604 | Credits | 3 | | Class | | VI semester | |
| Course Structure | TLP | Credits | Contact Hours | Work Load | 13Hrs/ Semester | | Assessment in Weightage | |
| | Theory | 3 | 3 | 3 | | | | |
| | Practice | - | - | - | Theory | Practical | CIE | SEE |
| | Tutorial | - | - | - | | | | |
| | Total | 3 | 3 | 3 | 39 | 0 | 50% | 50% |

COURSE OVERVIEW

The intent of the course is to give introduction to Natural Language Processing (NLP, a.k.a. computational linguistics), the 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 as listed in the course syllabus, algorithms for effectively solving these problems, and methods for evaluating their performance. There will be a focus on statistical algorithms to acquire the knowledge needed to perform language processing.

COURSE OBJECTIVE (S):

The objectives of this course are to:

1. Explain the fundamentals of natural language processing and python.
2. Discuss how to access the text corpora and Lexical Resources.
3. Demonstrate the writing the structured programs to process the raw text.
4. Describe role of Classifiers in Text processing.

COURSE OUTCOMES (COs)

After the completion of the course, the student will be able to:

| CO# | Course Outcomes | POs | PSOs |
|-----|-----------------------------------------------------------------------------------------------------------|--------|------|
| CO1 | Apply the concepts of Python to implement NLTK tool to solve a given real world problem. | 1 to 5 | 1,3 |
| CO2 | Develop an algorithm to access the text corpora and Lexical Resources to process the raw text. | 1 to 5 | 2 |
| CO3 | Make use of Categorizing and Tagging concepts to solve the given real-world problems. | 1 to 5 | 2,3 |
| CO4 | Analyze the performance of different classifiers in Text processing and Modelling Linguistic. Patterns | 1 to 5 | 1 |

| | | | |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|-----|
| CO5 | Learn new tools and technologies in the natural language processing and apply for suitable application development. | 1 to 5, 12 | 1,2 |
| CO6 | Develop solutions in the language processing to the complex problems, either individually or as a part of the team and report the results with proper analysis and interpretation. | 1 to 5, 9, 10 | 2,3 |

BLOOM'S LEVEL OF THE COURSE OUTCOMES

| CO# | Bloom's Level | | | | | |
|-----|------------------|--------------------|---------------|-----------------|------------------|----------------|
| | Remember (L1) | Understand (L2) | Apply (L3) | Analyze (L4) | Evaluate (L5) | Create (L6) |
| CO1 | | | √ | | | |
| CO2 | | | √ | | | |
| CO3 | | | √ | | | |
| CO4 | | | | √ | | |
| CO5 | | | √ | √ | | |
| CO6 | | | | √ | | |

COURSE ARTICULATION MATRIX

| CO#/ POs | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 | PSO3 |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|
| CO1 | 2 | 3 | 3 | 3 | 2 | | | | | | | | 3 | | 3 |
| CO2 | 2 | 3 | 2 | 3 | 3 | | | | | | | | | 3 | |
| CO3 | 2 | 3 | 2 | 3 | 2 | | | | | | | | | 3 | 3 |
| CO4 | 3 | 3 | 1 | 3 | 2 | | | | | | | | 3 | | |
| CO5 | 2 | 2 | 2 | 2 | 2 | | | | | | | 3 | 3 | 3 | |
| CO6 | 2 | 3 | 3 | 2 | 2 | | | | 2 | 2 | | | | 3 | 3 |

Note: 1-Low, 2-Medium, 3-High

COURSE CONTENT

THEORY:

UNIT-1

Language Processing and python, Accessing Text corpora and Lexical Analysis: Computing with language- Texts and words, a closer look at python: texts as list of words, computing with language: simple statistics, Automatic

natural language understanding; Accessing Text Corpora, Conditional Frequency Distributions, Lexical Resources, WordNet, Introduction to NLTK Tool. **Processing Raw Text:** Accessing Text from the Web and from Disk, Strings: 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 for Tokenizing Text, Segmentation, Formatting: From Lists to Strings.

UNIT-2

Categorizing and Tagging words: Using a Tagger, using a Tagger, Mapping Words to Properties Using Python Dictionaries, Automatic Tagging, N-Gram Tagging, Transformation-Based Tagging, How to Determine the Category of a Word. **Classifying Text:** Supervised Classification: Examples, Evaluation; Decision Trees, Naive Bayes Classifiers, Maximum Entropy Classifiers, Modelling Linguistic Patterns.

UNIT-3

Text Summarization: Text Summarization and Information Extraction, Feature Matrix, Single Value Decomposition, Keyphrase Extraction, Topic Modelling, Automated Document Summarization. **Text Similarity and Clustering:** Information Retrieval, Feature Engineering, Text Similarity, Analyzing Term Similarity, Analyzing Document Similarity, Document Clustering, Clustering Greatest Movies of All Time

UNIT-4

Semantic and Sentiment Analysis: Semantic Analysis, Exploring WordNet, Word Sense Disambiguation, Named Entity Recognition, Analyzing Semantic Representation, Sentiment Analysis

TEXT BOOKS:

1. Steven Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python", First Edition, O'Reilly Media, 2009.
2. Yuxi (Hayden) Liu, "Python Machine Learning by Example," First edition, Packt publisher, 2017.
3. Dipanjan Sarkar, "Text Analytics with Python A Practical real-world approach to Gaining Actionable Insights from Your Data", APress, 2016.

REFERENCE BOOKS:

1. James Allen, "Natural Language Understanding", Benjamin-Cummings Publishing Co., Inc. Redwood City, CA, USA, 1995.
2. Christopher D. Manning and Hinrich Schütze, "Foundations of Statistical Natural Language Processing", The MIT Press, 1999.
3. Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik, "A Comprehensive Grammar of English Language", Cambridge University Press, 1987.

JOURNALS/MAGAZINES

1. ACM Transactions on Language Processing
2. Elsevier Journal of cognitive systems research

SWAYAM/NPTEL/MOOCs:

1. <https://nptel.ac.in/courses/106/105/106105158/>
2. <https://nptel.ac.in/courses/106/106/106106211/>

Self-Learning Exercises:

Extracting information from Text, Exploring the 20 Newsgroups with Text Analysis Algorithms, Stock Price prediction with Regression Algorithms,

Best Practices:

- i) Data preparation stage
- ii) Training sets generation stage
- iii) Model training, evaluation and selection stage