PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada-520010
Re-accredited at 'A+'by the NAAC

Offered to: M.Sc. (Computer Science)

CourseName	Natural Language Processing	**L**	**T**	**P**	**C**	**CIA**	**SEE**	**TM**
CourseCode	22CS4E7	4	0	0	4	30	70	100
Year of Introduction: 2022	**Year of Offering:** 2022	**Year of Revision**: Nil	**Percentage of Revision:**					
L-Lecture, T-Tutorial, **P**-Practical, **C**-Credits, **CIA**-InternalMarks, **SEE**-ExternalMarks, **TM**-TotalMarks								

Course Description and Purpose: Natural Language Processing is a course that illustrates concepts of Understanding the Structure of a Sentences, Preprocessing, Feature Engineering and NLP Algorithms, Basic Feature Extraction Methods, Text Classifier, Text Summarization and Text Generation, Vector Representation.

.

Course Objectives: This course will help enable the students to understand and familiar with Understanding the Structure of a Sentences, Preprocessing, Feature Engineering and NLP Algorithms, Basic Feature Extraction Methods, Text classifier, Text Summarization and Text Generation, Vector Representation.

Course Outcomes: On successful completion students should be able to
CO1: Define Natural Language, NLP techniques, components of NLP to process basic text analytics.
CO2: Illustrate feature engineering strategies, Feature Extraction Methods for text data in Python.
CO3: Develop text summarization and generation models using NLP algorithms.
CO4: Analyze web scraping, data collection, and vector representation for text retrieval.
CO5:Evaluate sentiment analysis techniques and tools for text data interpretation

CO-PO MATRIX							
COURSE CODE	**CO-PO**	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**
	CO1	M				M	
	CO2	H		H			
	CO3	H		H			
	CO4	M		M			
	CO5	H				H	

UNIT-I (12 Hours)

Introduction: Understanding natural language processing- What is Natural Language?, What is Natural Language Processing**?,** Understanding basic application**s-** Understanding advanced applications, Advantages of togetherness NLP and Python, Text Analytics and NLP - Basic Text Analytics ,Various steps in NLP-Tokenization, PoS Tagging Removal, Normalization, Spelling, Stemming, Lemmatization, NER, Word Sense Disambiguation, Sentence Boundary Detection

UNIT-II (12 Hours)

Understanding the Structure of a Sentences: Understanding the components of NLP-NLU and NLG, Differences of NLU and NLG ,branches of NLP, What is context-free grammar?, Morphological analysis, Lexical analysis, Syntactic analysis, Semantic analysis.

Preprocessing:-Basic preprocessing ,Regular expressions, Basic level regular expression -Basic flags, Advanced level regular expression -Positive look ahead , Positive look behind ,Negative look ahead Negative look behind.

Feature Engineering and NLP Algorithms: What is feature engineering? What is the purpose of feature engineering?Basic feature of NLP -Parsers and parsing, Understanding the basics of parsers ,Understanding the concept of parsing ,Developing a parser from scratch- Types of grammar -Context-free grammar ,Probabilistic context-free grammar -Calculating the probability of a tree, Calculating the probability of a string

UNIT-III (12 Hours)

Basic Feature Extraction Methods: Introduction, Types of Data- Categorizing Data Based on Structure, Categorization of Data Based on Content, Cleaning Text Data-Tokenization, Types of Tokenizers, Issues with Tokenization, Stemming, Regexp Stemmer, The Porter Stemmer, Lemmatization, Language Translation, Stop Word Removal, Feature Extraction from Texts-Extracting General Features from Raw Text, Bag of Words ,TF-IDF, Feature Engineering- Word Clouds, Other Visualizations

UNIT-IV (12 Hours)

Collecting Text Data from the Web: Introduction, Collecting Data by Scraping Web Pages-Extraction of Tag-Based Information from HTML Files, Requesting Content from Web Pages-Collecting Online Text Data, Analyzing the Content of Jupyter Notebooks (in HTML Format), Extracting Information from an Online HTML Page, Dealing with Semi-Structured Data- Dealing with JSON Files, Dealing with a Local XML File

Text Summarization and Text Generation: Introduction, What is Automated Text Summarization?-Benefits of Automated Text Summarization, High-Level View of Text Summarization-Purpose, Input, Output, Extractive Text Summarization, Abstractive Text Summarization, Sequence to Sequence, Encoder Decoder, Summarizing Text Using Word Frequency-Word Frequency Text Summarization,

UNIIT-V (12 Hours)

Vector Representation: Introduction, Vector Definition**,** Why Vector Representations?**-**Encoding, Character-Level Encoding-Character Encoding Using ASCII Values, Character Encoding with the Help of NumPy Arrays, Positional Character-Level Encoding- Character-Level Encoding Using Positions**,** One-Hot Encoding-Key Steps in One-Hot Encoding, Character One-Hot Encoding – Manual.

Sentiment Analysis: Why is Sentiment Analysis Required?**,** Types of Sentiments**,** Applications of Sentiment Analysis**,** Tools Used for Sentiment Analysis,TextBlob**-**Basic Sentiment Analysis Using the TextBlob Library.

Prescribed Text Book			
	Author	Title	Publisher
1	Jalaj Thanaki	Python Natural Language Processing	Packt Publishing Ltd Ist. Edition 2017 UNIT-I,II
2	Sohom Gosh	Natural Language Processing Fundamentals	Packt Publishing Ltd. 1st Edition 2019 UNIT I ,II -III,IV and V

Reference Text Books			
	Author	Title	Publisher
1	Daniel Jurafsky, James H. Martin	Speech and Language Processing	Pearson 3rd edition 2021
2	Christopher D. Manning, Hinrich Schütze	Foundations of Statistical Natural Language Processing	The MIT Press, 1st edition 1999

PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada-520010
Re-accredited at 'A+'by the NAAC

M.Sc.(Computer Science) **Semester :IV**

Course Code: 22CS4E7 Course Name: Natural Language Processing

Time: 3 Hours **Max Marks: 70**

SECTION-A

Answer the following questions. (5×4=20Marks)

1. (a)Define Natural Language Processing.What are the advantages of NLP and Python?(CO1,L1)
(or)
(b) What are the basic applications of NLP.(CO1,L1)
2. (a) What are the differences between NLU and NLG?(CO1,L1)
(or)
(b) Define Regular expression. Explain basic regular expressions?(CO1,L1)
3. (a) Explain Types of Data used in Feature Extraction Method. (CO2,L2)
(or)
(b) Explain about porter stemmer. (CO2,L2)
4. (a)Explain Automated Text Summarizationand its benefits.(CO3,L2)
(or)
(b) Explain Collecting Data by Scraping Web Pages with example. (CO4,L2)
5. (a) Explain Character Encoding Using ASCII Values. (CO4,L2)
(or)
(b) Explain types of Sentiment Analysis.(CO5,L2)

SECTION-B

Answer the following questions. (5×10=50Marks)

6. (a) Define Natural Language.What are the Advanced Applications used in NLP?.(CO1,L1)
(or)
(b) Define Tokenization and PoS Tagging in NLP with example.(CO1,L1)

7. (a) Explain about Advanced Regular Expressions with example. (CO1,L5)
(or)
(b) Explain about CFG and PCFGs with examples. (CO1,L5)

8. (a) Explain about types of Tokenizers and issues with Tokenization. (CO2,L2)
(or)
(b) Explain about Feature Engineering. (CO2,L2)

9. (a)Explain Semi-Structured Data using XML and JSON files.(CO4,L5)
(or)
(b) Explain High-Level View of Text Summarization. (CO3,L5)

10.(a) Elaborate about one hot encoding with example? (CO5,L6)
(or)
(b) Develop Basic Sentiment Analysis using TextBlob library.(CO5,L6)