|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Course Code: CSE3015** | **Course Title: Natural Language Processing** | **TPC** | | **3** | **2** | **4** |
| **Version No.** | **2.1** | | | | | |
| **Course Pre-requisites/ Co-requisites** | **Introduction to Machine Learning (CSE3008)** | | | | | |
| **Anti-requisites (if any).** | **None** | | | | | |
| **Objectives:** | 1. To introduce students the fundamentals of Natural Language Processing. 2. Learn the techniques in natural language processing. 3. Be familiar with the natural language generation. 4. Be exposed to Text Mining. 5. Understand the information retrieval techniques | | | | | |
| **Expected Outcome:** | On completion of the course, students will have the ability to   1. Understand and implement word and syntactic level anlysis 2. Extract relation in text. 3. Implement the Python and NLTK libraries 4. Implementing Sematic analysis 5. Implement information retrievals | | | | | |
| **Module No. 1** | **Introduction to NLP** | | **6 Hours** | | | |
| Overview: Origins and challenges of NLP-Need of NLP, python and NLTK for NLP, Text Wrangling and cleansing- Text cleansing, sentence splitter, tokenization, stemming, lemmatization, stop word removal, rare word removal, spell correction | | | | | | |
| **Module No. 2** | **Parts of Speech Tagging** | | **6 Hours** | | | |
| Parts of Speech Tagging – Tagging in NLP, Sequential tagger, N-gram tagger, Regex tagger,Brill tagger, Machine learning taggers-MEC,HMM,CRF , NER tagger, | | | | | | |
| **Module No. 3** | **Parsing Structure in Text** | | **12 Hours** | | | |
| Shallow vs Deep parsing, Approaches in parsing, Types of parsing, Regex parser, Dependency parser, chunking, Information extraction, Relation Extraction, Building first NLP Application, Machine translation application. | | | | | | |
| **Module No. 4** | **Text Classification** | | **6 Hours** | | | |
| Types of learning techniques, Text Classification-Sampling, Naïve Bayes, Decision trees, Stochastic gradient descent, Support vector machine, Text clustering, | | | | | | |
| **Module No. 5** | **Web Crawling and Social Media Mining** | | **6 Hours** | | | |
| Web crawler – Writing first crawler – Data flow in Scrapy – Scrapy shell. Social Media Mining-Data Collections, Data Extraction, Geo visualization | | | | | | |
| **Module No. 6** | **Applications of NLP** | | **9 Hours** | | | |
| Applications of NLP: Transforming text, Sentiment Analysis, Information retrieval, text summarization, Question and Answering, Automatic Summarization | | | | | | |
| **Text Books**   1. Nitin Hardeniya, Jacob Perkins, Deepti Chopra, Nisheeth Joshi, Iti Mathur, “Natural Language Processing: Python and NLTK”, Packt publisher, 2016. | | | | | | |
| **References**   1. Steven Bird, ‎Ewan Klein, ‎Edward Loper, “Natural Language Processing with Python”, O’Reilly, 1st Edition 2009. 2. Jacob Perkins, “Python 3 Text Processing with NLTK 3 Cookbook”, Pearson Education, Second Edition, 2014. 3. Deepti Chopra, Nisheeth Joshi, Iti Mathur, “ Mastering Natural Language Processing with Python”, Packt **,** 2016. | | | | | | |
| **Lab Exercises**  1. Read the paragraph and obtain the frequency of words.  2. Read the content from a web page and extract the tokens /expression/word/number.  3. Read only the word content from webpage and display their frequency?  4.Plot the frequency for count obtained in question 3.  5.Write a program to slit sentences in a document?  6. Perform tokenizing and stemming by reading the input string?  7. Remove the stopwords and rareword in the document?  8. Identify the parts of speech in the document?  9. Implement the N-gram tagger?  10.Implement Regex tagger?  11. Implement Brill tagger?  12. Implement Maximum Entropy Classifier?  13.Implement NER tagger?  14. write a tagger that tags Date and Money expressions?  15. Define a grammar and obtain the sentences from the grammar?  16. Implement Regex parser?  17. Implement chunking using Shallow parsing?  18.Obtain the Named entity relations in the document?  19. Choose any news article with only contents of the news dumped into a text file and obtain the top 10 sentences?  20. Implement a text classification application?  21. Implement a text clustering application?  22. Implement a web crawler?  23. write an application for social media mining?  24. Identify the Facebook influencer ?  25. Visualize the influencer obtained in question 24? | | | | | | |
| **Mode of Evaluation** | Cumulative Lab Assessment 20%  Continuous Assessment Test-1 20%  Continuous Assessment Test-2 20%  Final Assessment Test 20%  Practical Assessment (Mini Project) 20% | | | | | |
| **Recommended by the Board of Studies on** | 12th BoS, 29.04.2023 | | | | | |
| **Date of Approval by the Academic Council** | 10th AC (01.06.2023) | | | | | |