**MINOR PROJECT**

**ON**

**“TEXT SUMMARIZATION**

**USING NLP IN PYTHON”**

**SUBMITTED BY**

**NAG ASHISH S.V**

**BATCH 2-ML AND PYTHON**

**(gmailid:ashishsv028@gmail.com)**

**SUBMITTED TO**

**VERZEO EDUTECH**

**During The Months**

**June-July 2019**

**LIBRARIES AND TOOLS USED IN THE PROJECT:**

-----------------------------------------------

Language used: Python 3.6

Editor used: Pycharm 2017-3.3

Libraries used: Sumy and NLTK

ABOUT THE PROJECT:

**1)Install the package Sumy and NLTK.**

**2)Import the below packages from the above libraries:**

**a) from** sumy.parsers.plaintext **import** PlaintextParser

**b) from** sumy.nlp.tokenizers **import** Tokenizer

**c) from** sumy.summarizers.lex\_rank **import** LexRankSummarizer

**3)Read the file “mldoc.txt” containing the required text(paragraph).**

**4)Parse the input para by tokenizing it using**

**the function PlaintextParser.from\_sting().**

**5)Use the LexRank Algorithm to summarize the paragraph.**

**6)create the object of LexRankSummarizer().**

**7)call the summarizer function with the following arguments.**

**a)parser.document(parser object)**

**b)Number of lines required in summary.**

**8)print the sentences from the summary,which are**

**the required summary lines.**

**ABOUT LEXRANK ALGORITHM:**

**It is an unsupervised approach to text summarization based on graph-based centrality scoring of sentences.**

**The main idea is that sentence recommend other similar sentences to the reader.**

**Thus if one sentence is very similar to many others, it will likely be a sentence of great importance.**

**OTHER ALGORITHMS FOR SUMMARIZATION OFFERED BY SUMY LIBRARY:**

* Luhn – heurestic method.
* Latent Semantic Analysis(LSA).
* Edmundson heurestic method with previous statistic research.
* LexRank – Unsupervised approach inspired by algorithms PageRank and HITS.
* TextRank.
* SumBasic – Method that is often used as a baseline in the literature.
* KL-Sum – Method that greedily adds sentences to a summary so long as it decreases the KL Divergence.

**CONCLUSION:**

**The overall summarization is highly accurate and only important sentences are picked for the summary making the summary more readable for the users. The LexRank algorithm worked precisely for summarizing the text.**