**MINI PROJECT – II**

**(2020-21)**

**Text Summarizer**

**SYNOPSIS**



**Institute of Engineering & Technology**

***Made By***

Afreen Hasan

(171500022)

## Supervised By

Mr. Vinay Agrawal

**Assistant Professor**

**Department of Computer Engineering & Applications**

**About the Project:**

Automatic text summarization is basically summarizing of the given paragraph using natural language processing and machine learning. There has been an explosion in the amount of text data from a variety of sources. This volume of text is an invaluable source of information and knowledge which needs to be effectively summarized to be useful. In this review, the main approaches to automatic text summarization are described. We review the different processes for summarization and describe the effectiveness and shortcomings of the different methods. Two types will be used i.e.-extractive approach and abstractive approach. The basic idea behind summarization is finding the subset of the data which contains the information of all the set. There is a great need to reduce unnecessary data. It is very difficult to summarize the document manually so there is the great need of automatic methods. Approaches have been proposed inspired by the application of deep learning methods for automatic machine translation, specifically by framing the problem of text summarization as a sequence-to-sequence learning problem.

**Problem Statement:**

In the new period, where tremendous measure of data is accessible on the Web, it is most vital to give the enhanced gadget to get data rapidly. It is extremely intense for individuals to physically pick the synopsis of expansive archives of content. So there is an issue of scanning for vital reports from the accessible archives and discovering essential data. Along these lines programmed content rundown is the need of great importance. Content rundown is the way toward recognizing the most vital important data in a record or set of related archives. What’s more, compact them into a shorter rendition looking after its implications.

**Motivation:**

The objective of the project is to understand the concepts of natural language processing and creating a tool for text summarization. The concern in automatic summarization is increasing broadly so the manual work is removed. The project concentrates creating a tool which automatically summarizes the document.

**Scope:**

The project is wide in scope all of the limitations stated below may seem to contradict that, but they are the only restrictions applied. This project looks at single document summarization the area of multi document summarization is not covered. Also, the summaries produced are largely extracts of the document being summarized, rather than newly generated abstracts. The parameters used are optimal for news articles, although that can be changed easily. With regard to language simplification, only lexical changes were considered syntactic changes were not. Background information was limited to biographical information and maps.

**Future Prospects:**

We have implemented Automatic text summarization using abstractive method. Further, after using RNN and LSTM the accuracy is still very low for summarizer. Furthermore, we will be using machine learning for semantic text summarization for more accurate summaries and will try to make a grader which will grade the document according to English grammar. There are many text summarizers available but all does not give appropriate result. Thus we will be using machine learning algorithm to increase the effectiveness of the automatic summarizer.

**Requirements:**

* **Hardware Requirements:**

1. Cloud
2. Computer with GPU capabilities

* **Software Requirements:**

1. Python environment

* **Functional Requirements;**

1. Natural Language Processing
2. Python Library Nltk