

# TEXT SUMMARIZATION TOOL USING PYTHON

<sup>1</sup>Aniketh Jadhav, <sup>2</sup>Ninad Ghandat, <sup>3</sup>Shubham Pawar

<sup>1, 2, 3</sup> Student,

Department Of Information Technology, MET Bhujbal Knowledge City Institute of Engineering, Nashik, Maharashtra, India.

*Abstract:* In recent years, an enormous amount of textual data, from various sources are emerging day-by-day. This data contains essential information and knowledge that needs to be effectively summarized to be useful. Natural Language Processing (NLP) is used to parse or process the article or text file. Polarity analyzes each and every word from the article, and calculates a number, and this number will specify the sentiment of the article has been given as input (positive or negative). Reading long texts and articles is not possible for everyone in this smart and fast day-to-day life. The tool will be made with the help of Natural Language Toolkit. This tool aims to deliver short yet meaningful summaries of text documents to the end users.

# Index Terms - Summarization, Polarity, Natural Language Toolkit, Natural Language Processing

# I. INTRODUCTION

In today's modern day-to-day life, we come across vast number of texts, articles, documents which we are not able to read and understand at all the times. Reading summaries with semantic information is the best way to overcome this issue. A summary has all the semantic meaning of the enormous size of texts or articles, and it is small in size. A person can refer to a summary rather than reading big texts and articles to obtain whatever information they want. The need for a good Text Summarization Tool is to reduce and save time by creating significant and shorts summaries. This could help any individual in writing better reports, finding more relevant information in less time. Text summarization tool can be also used for Information Retrieval purposes. A good summary should also consist of other aspects such as covering the whole topic, relevancy, and readability. This tool can be used by any personnel, for any activity. This literature paper aims to contribute to the existing literature on text summarization techniques, and its applications.

The exponential growth of digital content across various domains, such as news articles, research papers, online reviews, and social media posts, has resulted in an information overload for users. Text summarization tool will serve as an indispensable tool for addressing this challenge by automatically generating condensed and precise summaries of the original text, enabling users to quickly grasp the core essence without the need for difficult reading. By reducing information redundancy and providing concise summaries, text summarization tool empowers users to save time, make informed decisions, and efficiently navigate through vast amounts of textual data.

Generally, there are three types of summarization techniques:

- **1**) Abstractive summarization
- 2) Extractive summarization
- 3) Hybrid summarization

# **II. LITERATURE REVIEW**

G. Ramesh, Ravali Boorugu - "A Survey on NLP Based Text Summarization for Summarizing Product Reviews" (2020) [1] - This paper is a survey on the various types of text summarization techniques starting from the basic to the advanced techniques. Narendra Andhale - "An overview of Text Summarization Techniques" (2016) [2] – Text summarization approaches are classified into two categories: Extractive and Abstractive.

P. Janjanam, CH Pradeep Reddy - "Text Summarization: An Essential Study" (2019) [3] - This survey is intended to make an extensive study from features representation to sentence selection and summary generation using machine learning.

J.N. Madhuri, R. Ganesh Kumar – "Extractive Text Summarization Using Sentence Ranking" (2019) [4] - In this paper, a statistical method to perform an extractive text summarization on single document is demonstrated.

IJNRD2305581

#### © 2023 IJNRD | Volume 8, Issue 5 May 2023 | ISSN: 2456-4184 | IJNRD.ORG

M. Patel, Satyadev Vyas, K. Maurya – "Machine Learning Approach for Automatic Text Summarization Using Neural Networks" – 2018 [5] – In this paper, the authors have explained use of Machine Learning and Deep Learning, and the neural network libraries used in Text Summarization process.

## **III. NEED OF THE PROPOSED TOOL**

The need for text summarization tool arises from the continuously increasing volume of information available in today's digital era. With a constant increase of textual data from sources such as news articles, research papers, social media posts, and online reviews, individuals and organizations face the challenge of efficiently processing and extracting valuable insights from these vast amounts of text. A good text summarization tool is necessary due to the following reasons:

- 1. Information Overload: The exponential growth of digital content has led to information overload, where individuals are inundated with more information than they can effectively process. Text summarization addresses this issue by condensing lengthy documents into concise summaries, enabling users to quickly grasp the main ideas and key points without having to read through the entire text.
- 2. Time Efficiency: In today's fast-paced world, time is a valuable resource. Text summarization allows users to save time by providing them with small versions of texts. Instead of spending significant time reading lengthy articles or documents, users can efficiently scan summaries to obtain the most relevant information, which helps individuals in quicker decision-making and task completion.
- 3. Decision-Making Support: Text summarization helps decision-makers in various fields by providing them with a comprehensive overview of the information they need to consider. Whether it's executives making strategic decisions or researchers analyzing scientific literature, text summarization helps distill the essential elements and enables informed decision-making.
- 4. Information Extraction: Text summarization algorithms excel at identifying important entities, key phrases, and relevant concepts within a text. By extracting these elements and presenting them in a concise format, summarization techniques assist in knowledge extraction, information retrieval, and content analysis tasks.

# IV. RECENT STUDIES AND SUMMARIZATION APPROACHES

Text summarization techniques can be broadly categorized into two main types: extractive summarization and abstractive summarization. Both approaches aim to condense the original text into a shorter, more concise form, but they differ in their methods of generating summaries. The types of text summarization are as following:

- 1. Extractive Summarization: Extractive summarization involves selecting and assembling the most important sentences, phrases, or passages from the source text to form the summary. This approach extracts existing sentences verbatim from the original text and does not generate new sentences.
- 2. Abstractive Summarization: Abstractive summarization involves generating new sentences that capture the essence of the original text while using different wording and phrasing. This approach aims to produce summaries that seem to be generated by humans rather than by machines and AI.
- 3. Hybrid Summarization: It is the combination of both, Extractive and Abstractive Summarization techniques.

Extractive summarization selects and concatenates important sentences from the source text, whereas abstractive summarization generates new sentences that capture the essence of the original text. All three approaches have their pros and cons, and researchers continue to explore and improve these methods to enhance the quality and effectiveness of text summarization systems.



## V. PROPOSED SYSTEM

The user can either select the text or the URL of an article from any website (example: Wikipedia.) This article is pasted into input section of the Text Summarization Tool GUI. The tool will download the data found on the webpage of the entered URL. The Natural Language Processing (NLP) is used for parsing or processing of the texts and articles. The data obtained from URL is then summarized. Sentiment analysis of that text or article is calculated and displayed along with the final summary. Polarity analyses every word in the text and calculates a number which denotes the sentiment of that particular document.

The input will be in the form of an URL of that respective article or text document. The tool will start with the processing of the contents in the article. The processing of the web contents will be carried out by NLP. The feature extraction will be done by the

#### © 2023 IJNRD | Volume 8, Issue 5 May 2023 | ISSN: 2456-4184 | IJNRD.ORG

'newspaper' library. The newspaper library will extract the useful information from the web page and discard the irrelevant information. The TextBlob library will do the sentimental analysis of the web page contents, along with the calculation of polarity. With the help of these processes, a long text or article can be converted into a small yet meaningful, precise summary.



#### VI. METHODOLOGY

[1] Natural Language Toolkit (NLTK): It is a platform which is used for building and implementing Python programs to work with human language data. Some of the operations of text processing include classification, tokenization, stemming, ranking, etc. [2] Natural Language Processing (NLP): NLP is a subfield of artificial intelligence. It mainly focuses on the interaction between human language and the computers. It helps in the development of algorithms and models that enable computers to understand, interpret, and generate human language in a meaningful way. [3] Sentiment Analysis: It is a technique used to determine the sentiment or emotion expressed in a piece of text. It involves analyzing textual data to classify it as positive, negative, or neutral, which indicates the writer's attitude or opinion towards a particular subject. [4] Tkinter: It is a built-in graphical user interface (GUI) toolkit in Python. It provides a set of tools and widgets that allow developers to create desktop applications with graphical interfaces. [5] Polarity: Polarity aims to differentiate between the opinion into positive and negative. It lies between the range of [-1, 1], where 1 means positive statement and -1 means negative statement. Examples: the words like 'good', 'amazing', 'perfect' are positive meaning words. While, the words like 'terrible', 'awful' are negative words. From these words, a Polarity score is calculated by the Sentiment Analysis, which tells overall that if the article or text is positive or negative.

#### **VII. APPLICATIONS**

[1] Chatbots and Virtual Assistants: Text summarization helps chatbots and virtual assistants understand and respond to user queries by extracting relevant information from large text sources. [2] Books and Literature: A general overview or preface of a book contains information related to the content of that particular book, which can help a customer to sort which book suits them. [3] E-learning and Teaching: reading big paragraphs is hectic as well as time consuming for students. Summaries of particular topics can be read by students which are easy to understand, to get a quick idea about that topic. Teachers can refer to precise summaries rather than reading big articles and wasting their valuable time. [4] E-mail Filtering: Summarizing emails or categorizing them based on their content can help users in prioritizing and managing large volumes of inbox, unread messages.

#### VIII. RESULTS

#### Input:

The user is intended to enter the required text or the URL of that article. The data obtained from the webpage is analyzed and further converted into precise summaries using Python libraries. Example: <u>https://en.wikipedia.org/wiki/2022\_FIFA\_World\_Cup/</u>

#### **Expected Output:**

Summary: The summary processed with the help of libraries of Python will be displayed to the user. Polarity of the article: The polarity calculated will be displayed indicating the sentiment of the text.

#### **IX. CONCLUSION**

Text summarization tool is an important resource that can assist in easy and quick extraction of the key information and main ideas from a lengthy text. It saves time and effort by automatically shrinking huge documents into shorter summaries. The tool is helpful where individuals may want to grasp the main points before deciding to invest time in reading the complete text. Text summarization is useful for Natural Language Processing tasks (NLP) like Answering Questions or Text Classification, and other computer science fields like Information Retrieval. The access time for information searching can also be improved with the help of the proposed tool.

IJNRD2305581

# REFERENCES

[1] Ravali Boorugu, G. Ramesh – "A Survey on NLP Based Text Summarization for Summarizing Product Reviews" – 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)

[2] N. Andhale – "An overview of Text Summarization Techniques" – 2016 International Conference on Computing Communication Control and automation (ICCUBEA)

[3] P. Janjanam, CH Pradeep Reddy – "Text Summarization: An Essential Study" – 2019 International Conference on Computational Intelligence in Data Science (ICCIDS)

[4] J.N. Madhuri, R. Ganesh Kumar – "Extractive Text Summarization Using Sentence Ranking" 2019 International Conference on Data Science and Communication (IconDSC)

[5] M. Patel, Satyadev Vyas, K. Maurya – "Machine Learning Approach for Automatic Text Summarization Using Neural Networks" – 2018 International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) [Vol. 7, Issue 1]

