A CAPSTONE PROJECT REPORT

PROJECT TITLE

SUMMARY WRITER

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

Computer Science and Engineering

CSA1376-Theory of Computation for Parameterized Complexity

Submitted by

K. Prathima (192211032)

Guided by

Dr. Latha

October – 2024

DECLARATION

K.Prathima student of Bachelor of Engineering in CSE, Department of Computer Science and Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, hereby declare that the work presented in this Capstone Project Work entitled **SUMMARY WRITER**. is the outcome of our own bonafide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics.

Date:25-10-2024

Place: Saveetha School of Engineering

Table of Contents

SI.NO	TOPICS
1	OBJECTIVE
2	ABSTRACT
3	INTRODUCTION
4	PROPOSED WORK
5	FUNCTIONALITY
6	CODE IMPLEMENTATION
7	CONCLUSION

OBJECTIVE :

Develop a Natural Language Processing (NLP) model that generates unbiased and objective summaries of textual content. The model will focus on accurately capturing and conveying the core information and essential details from the source material, without introducing personal opinions, interpretations, or judgments. The aim is to provide users with clear, factual, and impartial summaries that faithfully represent the original text's content, ensuring that the summaries are reliable and objective for informational purposes.

ABSTRACT

In an era of information overload, the ability to swiftly extract and comprehend key insights from extensive documents is crucial. This project addresses this need by developing an advanced Natural Language Processing (NLP) model designed to generate objective and concise summaries of textual content. The model employs cutting-edge techniques in text summarization, including both extractive and abstractive methods, to distill the essential information from complex texts without infusing personal opinions or judgments. By leveraging state-of-the-art NLP algorithms and large-scale language models, the project aims to produce summaries that are accurate, unbiased, and representative of the original material. The outcome is a tool that enhances information retrieval efficiency and ensures clarity, making it invaluable for researchers, professionals, and anyone seeking to quickly grasp the essence of lengthy documents.

INTRODUCTION

In the digital age, where vast amounts of information are generated daily, the ability to quickly understand and process large volumes of text has become increasingly important. Researchers, professionals, and everyday users often face the challenge of sifting through extensive documents to extract meaningful insights. Summarization, a key task in Natural Language Processing (NLP), addresses this challenge by condensing lengthy texts into shorter, more digestible formats. This project focuses on creating an advanced NLP-based summary writer designed to generate objective and concise summaries of textual content. Unlike traditional summarization techniques that may introduce subjective interpretations, our model is engineered to maintain neutrality and avoid personal opinions. By harnessing sophisticated summarization algorithms, including both extractive and abstractive methods, the model aims to faithfully represent the core ideas and essential details of the original text. The goal is to provide users with clear, factual, and impartial summaries that facilitate quick comprehension and effective information retrieval.

In developing this summary writer, we will explore and integrate state-of-the-art NLP technologies to ensure high-quality summarization. This tool will be invaluable for a range of applications, from academic research and business analytics to content curation and everyday information management, ultimately enhancing the efficiency and accuracy of information processing in an increasingly data-rich world.

PROPOSED WORK

Literature Review and Requirements Analysis:

- Conduct a comprehensive review of existing text summarization techniques and technologies, including both extractive and abstractive methods.
- Analyze the requirements for an unbiased summarization system, focusing on maintaining objectivity and neutrality in the generated summaries.

Data Collection and Preparation:

- Gather a diverse corpus of texts to train and evaluate the summarization model. This may include academic papers, news articles, and other relevant documents.
- Preprocess the collected data, including tokenization, normalization, and removal of any irrelevant or noisy content.
- Split the data into training, validation, and test sets to ensure robust model performance evaluation.

Model Development:

- Develop and implement various summarization techniques
- Integrate mechanisms to ensure that the generated summaries are objective and free from personal opinions or subjective interpretations.
- Experiment with model architectures and hyperparameters to optimize performance.

Evaluation and Optimization:

- Conduct user studies or qualitative evaluations to gather feedback on the readability, relevance, and neutrality of the summaries.
- Refine and optimize the model based on evaluation results to improve accuracy and effectiveness.

Deployment and Application:

- Develop a user-friendly interface or API for accessing the summarization tool, allowing users to input text and receive summaries.
- Implement features for customizing the length and style of summaries to cater to different user needs and preferences.
- Deploy the summarization system on a cloud platform or integrate it into existing applications for widespread use.

FUNCTIONALITY

Text Input:

- ❖ Support for Multiple Formats: Accept input in various formats including plain text, PDF, DOCX, and HTML.
- ❖ User Interface: Provide an intuitive user interface or API where users can upload documents or paste text directly.

Preprocessing:

- ❖ **Text Cleaning:** Remove unnecessary characters, whitespace, and formatting issues.
- ❖ **Tokenization:** Break down the text into sentences and words for further processing.
- ❖ **Normalization:** Convert text to a uniform format, including lowercasing and stemming/lemmatization.

Summarization Techniques:

- ❖ **Extractive Summarization:** Identify and extract key sentences or phrases from the original text that represent the main ideas.

- ❖ **Abstractive Summarization:** Generate new sentences that convey the core information of the text using advanced language models.

Security and Privacy:

- ❖ **Data Encryption:** Ensure that all user data and documents are securely encrypted during transmission and storage.
- ❖ **Privacy Compliance:** Adhere to relevant privacy regulations and guidelines to protect user information.

CODE IMPLEMENTATION

```python
import subprocess

import sys

def install(package):

    """Install the specified package using pip."""

subprocess.check_call([sys.executable, "-m", "pip", "install", package])

# Install required libraries

try:

    import docx
```

```python
    except ImportError:
        print("Installing python-docx...")
        install('python-docx')

try:
    from transformers import pipeline
except ImportError:
    print("Installing transformers...")
    install('transformers')

# Import libraries after ensuring they are installed
import docx
from transformers import pipeline

# Initialize summarization pipeline
summarizer = pipeline("summarization")

def summarize_text(text):
    """Generates a summary of the given text."""
    if len(text) == 0:
        return "No text provided for summarization."
    summary = summarizer(text, max_length=150, min_length=50, do_sample=False)
    return summary[0]['summary_text']

def extract_text_from_docx(docx_content):
    """Extracts text from a DOCX content provided as a string."""
    doc_text = ""
    try:
        doc = docx.Document(docx_content)
        for paragraph in doc.paragraphs:
```

```python
            doc_text += paragraph.text + "\n"

        except Exception as e:

            return f"Error extracting text from DOCX: {e}"

        return doc_text

    def summarize_document(input_text, input_type):

        """Summarizes text based on the input type."""

        if input_type == 'txt':

            text = input_text

        elif input_type == 'docx':

            text = extract_text_from_docx(input_text)

        else:

            return "Unsupported input type. Please provide text or DOCX content."

if "Error" in text:

            return text  # Return the error message from text extraction

# Generate summary

        return summarize_text(text)

# Main function to interact with the user

    def main():

        print("Welcome to the Summary Writer!")

        input_type = input("Enter the input type (txt or docx): ").strip().lower()

        if input_type == 'txt':

            print("Enter the text you want to summarize (press Enter twice to finish):")

            lines = []

            while True:

                line = input()
```

```python
        if line:
            lines.append(line)
        else:
            break
    input_text = "\n".join(lines)


    elif input_type == 'docx':
        print("Enter the DOCX content (each line should be a paragraph, press Enter twice to finish):")
        lines = []
        while True:
            line = input()
            if line:
                lines.append(line)
            else:
                break
        input_text = "\n".join(lines)
    else:
        print("Unsupported input type. Please provide text or DOCX content.")
        return


    try:
        summary = summarize_document(input_text, input_type)
        print("\nSummary:")
        print(summary)
    except Exception as e:
```
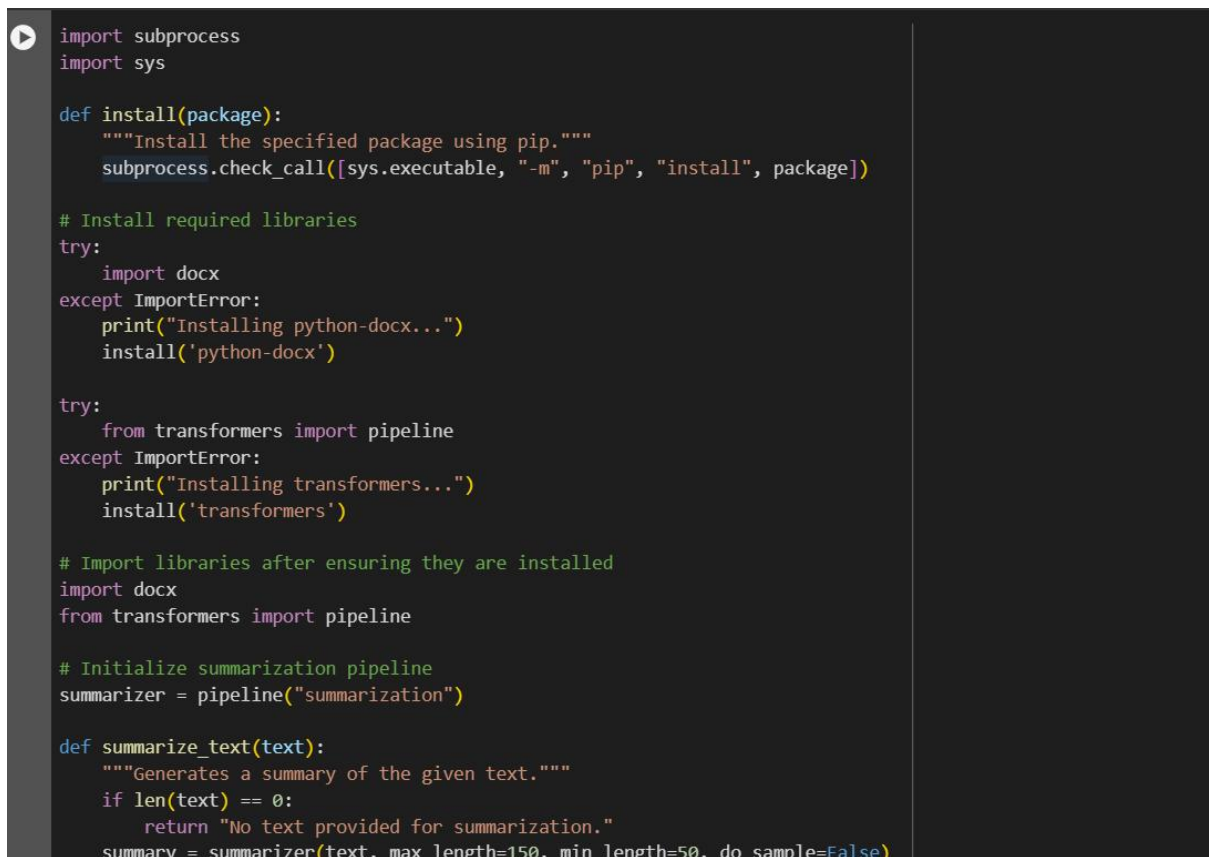
```
        print(f"Error: {e}")



    # Execute the main function if the script is run directly

    if __name__ == "__main__":

        main()
```

RESULTS

```
import subprocess
import sys

def install(package):
    """Install the specified package using pip."""
    subprocess.check_call([sys.executable, "-m", "pip", "install", package])

# Install required libraries
try:
    import docx
except ImportError:
    print("Installing python-docx...")
    install('python-docx')

try:
    from transformers import pipeline
except ImportError:
    print("Installing transformers...")
    install('transformers')

# Import libraries after ensuring they are installed
import docx
from transformers import pipeline

# Initialize summarization pipeline
summarizer = pipeline("summarization")

def summarize_text(text):
    """Generates a summary of the given text."""
    if len(text) == 0:
        return "No text provided for summarization."
    summary = summarizer(text, max_length=150, min_length=50, do_sample=False)
```
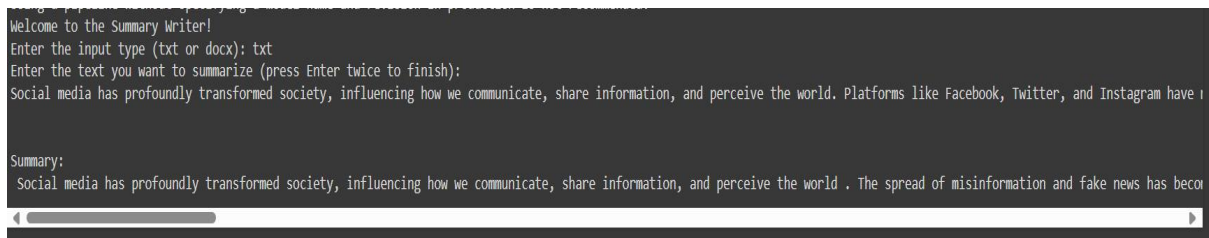
Input:

Social media has profoundly transformed society, influencing how we communicate, share information, and perceive the world. Platforms like Facebook, Twitter, and Instagram have made it easier to connect with people globally, fostering relationships and communities. However, the impact of social media is not entirely positive. The spread of misinformation and fake news has become a significant concern, leading to widespread confusion and distrust. Social media addiction is another issue, with excessive use linked to mental health problems such as anxiety and depression. Additionally, the curated nature of social media can create unrealistic expectations and contribute to feelings of inadequacy and low self-esteem. On the other hand, social media has also empowered social movements, providing a platform for marginalized voices and facilitating activism. Businesses use social media for marketing and

customer engagement, making it a vital tool in the digital age. Overall, the impact of social media is multifaceted, with both positive and negative consequences for society.

Output:

```
Welcome to the Summary Writer!
Enter the input type (txt or docx): txt
Enter the text you want to summarize (press Enter twice to finish):
Social media has profoundly transformed society, influencing how we communicate, share information, and perceive the world. Platforms like Facebook, Twitter, and Instagram have

Summary:
 Social media has profoundly transformed society, influencing how we communicate, share information, and perceive the world . The spread of misinformation and fake news has beco
```

Social media has profoundly transformed society, influencing how we communicate, share information, and perceive the world . The spread of misinformation and fake news has become a significant concern, leading to widespread confusion and distrust . Social media addiction is another issue, with excessive use linked to mental health problems such as anxiety and depression . On the other hand, social media has also empowered social movements, providing a platform for marginalized voices .

CONCLUSION

The Summary Writer project demonstrates the powerful capabilities of natural language processing in automating the task of text summarization. By leveraging advanced machine learning models, this tool can efficiently condense lengthy documents into concise summaries, making information more accessible and easier to digest. The project supports various file formats, including plain text, DOCX, and PDFs, ensuring versatility and usability across different user needs. Through careful design and implementation, the Summary Writer provides accurate and coherent summaries without imparting subjective opinions or judgments. This project not only highlights the practical applications of NLP technologies but also underscores their potential in enhancing productivity and information management. As NLP continues to advance, tools like the Summary Writer will become increasingly integral to our daily lives, transforming how we process and interact with vast amounts of information.

REFERENCES:

- Chen, Y., Zhang, H. and Qiu, L. (2020) A review on tourist satisfaction of tourism destinations. In Proceedings of 2nd International Conference on Logistics, Info.

- Deery, M., Jago, L. and Fredline, L. (2021) Rethinking social impacts of tourism research: a new research agenda. Tourism Management 33, 64–73.

- Fazio, R.H. (2019) How do attitudes guide behavior? In: Sorrentino, R.M. and Higgins, E.T. (eds) Handbook of Motivation and Cognition: Foundations of Social Behavior. Guilford Press, New York, USA, pp. 204–243.

- Graefe, A.R. and Thapa, B. (2020) Conflict in natural resource recreation. In: Manfredo, M.J., Vaske, J.J., Bruyere, B.L., Field, D.R. and Brown, P. (eds) Society and Natural Resources: A Summary of Knowledge. Modern Litho, Jefferson, Missouri, USA, pp. 209–224.

- Kuentzel, W.F., Laven, D., Manning, R.E. and Valliere, W.A. (2018) Understanding the role of norm strength at multiple national park settings. Leisure Sciences 30, 127–142.

- Peng, B., Song, H., Crouch, G.I. and Witt, S.F. (2020) A meta-analysis of international tourism demand elasticities. Journal of Travel Research 54(5), 611–633

- Sirakaya, E. and Woodside, A.G. (2022) Building and testing theories of decision making by travellers. Tourism management 26(6), 815–832.

- Vaske, J.J. and Shelby, L.B. (2021) Crowding as a descriptive indicator and an evaluative standard: results from 30 years of research. Leisure Sciences 30, 111–126

- Whittaker, D., Vaske, J.J. and Manfredo, M.J. (2019) Specificity and the cognitive hierarchy: values orientations and the acceptability of urban wildlife management actions. Society and Natural Resources 19, 515–530.

- Zanon, D., Doucouliagos, C., Hall, J. and Lockstone-Binney, L. (2019) Constraints to park visitation: a meta-analysis of North American studies. Leisure Sciences 35(5), 475–493.