Multilingual Video Summarization using NLP and Machine Learning

Dr. P. Rajeshwari[1], B. Manisha[2], R. Sandhya[3], Y. Vivek Reddy[4]

[1]CSE Department & ACE Engineering College
[2]CSE Department & ACE Engineering College
[3]CSE Department & ACE Engineering College
[4]CSE Department & ACE Engineering College

---***---

Abstract - Multilingual video summarization combines characteristic dialect handling (NLP) and machine learning to condense video substance into brief outlines over different dialects. This consider proposes a novel system leveraging robotized transcript extraction and summarization pipelines to create an English rundown, encourage deciphered into French, Spanish, and German. Utilizing progressed NLP models like Embracing Confront Transformers and YouTube Transcript API, our strategy productively forms recordings to create high-quality, important outlines within the target dialects. Comes about highlight the exactness and relevant pertinence of the interpretations, illustrating the system's potential in multilingual openness. The discoveries contribute to bridging etymological obstructions in video substance understanding and open roads for upgrading worldwide substance openness.

Key Words: video summarization, multilingual translation, NLP techniques.

1.INTRODUCTION

The developing mixed media substance on stages like YouTube highlights the require for instruments to upgrade worldwide availability. Video summarization, a subset of NLP, condenses long substance into brief outlines whereas protecting the center message. In spite of the fact that monolingual summarization has seen noteworthy headways, multilingual summarization remains underexplored.

This ponder proposes a system to summarize video substance in English and interpret it into different dialects utilizing state-of-the-art NLP models like Embracing Confront Transformers. The approach coordinating transcript extraction, summarization, and interpretation pipelines.

The inquire about hypothesizes that combining pre-trained summarization and interpretation models can create high-quality multilingual video outlines, progressing the field and progressing worldwide substance availability.

2. OBJECTIVES

Make a framework that forms video transcripts and produces brief, important English outlines whereas keeping up the center setting of the substance.

Decipher the produced English rundowns into numerous dialects (e.g., French, Spanish, and German) to cater to a worldwide gathering of people.

Use pre-trained models such as Embracing Confront Transformers and Marian MT to perform summarization and interpretation errands successfully and productively.

Assess the created rundowns and interpretations for exactness, coherence, and relevant significance utilizing measurements like BLEU scores and human input.

Contribute to decreasing dialect boundaries in interactive media substance, making it more comprehensive and open to non-English-speaking clients.

Survey the confinements of the current system, such as relevant mistakes and computational overhead, and recommend future bearings for optimization.

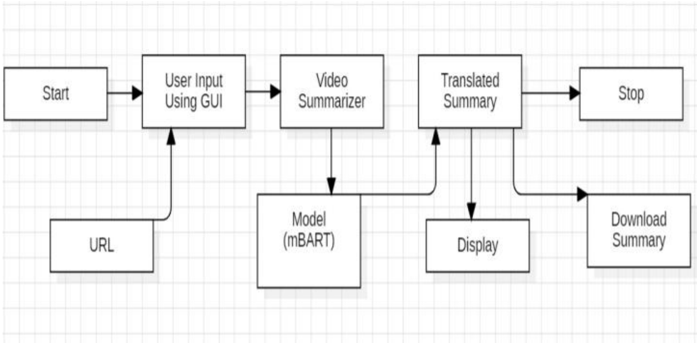
2.1 PROBLEM STATEMENT

With the exponential development of mixed media substance on stages like YouTube, getting to and understanding video substance over distinctive dialects has ended up a noteworthy challenge for a worldwide group of onlookers. Whereas progresses in Common Dialect Preparing (NLP) have moved forward monolingual summarization, multilingual video summarization remains underexplored. The need of proficient systems to create high-quality outlines and interpret them into different dialects limits substance openness and inclusivity for non-native speakers. This calls for a arrangement that addresses both summarization and multilingual interpretation successfully.

2.2 PROPOSED SYSTEM

The proposed framework points to address the challenges of multilingual video summarization by presenting a comprehensive system that coordinating progressed NLP procedures. The framework starts with transcript extraction utilizing Programmed Discourse Acknowledgment (ASR) instruments like Whisper or YouTube's captioning framework to get an exact literary representation of video substance. This is often taken after by content summarization, leveraging pre-trained models such as BART or T5 from Embracing Confront Transformers to condense the transcript into a brief and significant English rundown, protecting the center message. At long last, the summarized substance is interpreted into user-preferred dialects utilizing advanced interpretation models like Marian MT or Google Interpret API, guaranteeing syntactic precision and relevant pertinence. By combining these components, the framework empowers effective multilingual video summarization, improving worldwide substance availability and inclusivity.

2.3 SYSTEM ARCHITECTURE



3. METHODOLOGY

1) Input Video Interface

The method starts with the client giving a YouTube video interface as input. The essential objective of this step is to extricate the video ID from the URL, which serves as a reference to get the transcript of the video in ensuing steps. This guarantees that the method can handle energetic video inputs productively and plans the information pipeline for transcript recovery.

2) Transcript Extraction

In this step, the objective is to recover the video transcript in English utilizing the YouTube Transcript API. The video ID gotten within the past step is utilized to ask the transcript. Once recovered, the transcript content is collected and organized, planning it as input for the summarization module. This step lays the establishment for producing significant rundowns by organizing the crude transcript information successfully.

3) Content Summarization

The objective of this step is to produce a brief and significant outline of the video transcript. The Embracing Confront Transformer-based summarization demonstrate is utilized for this reason. In the event that the transcript surpasses the model's input estimate, it is partitioned into littler chunks. Rundowns are created for each chunk, and these are combined to form a comprehensive English rundown. To refine the comes about advance, a minimized outline is made, which condenses the substance indeed more whereas holding its basic meaning. This guarantees both nitty gritty and quick-reference rundowns are accessible.

4) Multilingual Interpretation

In this step, the objective is to create the summarized substance available in different dialects, such as French, Spanish, and German. Marian MT models from Embracing Confront are utilized for interpretation. The minimized English rundown is interpreted into the target dialects utilizing pre-trained interpretation models. This step guarantees that the substance is comprehensive and can cater to a assorted, multilingual group of onlookers.

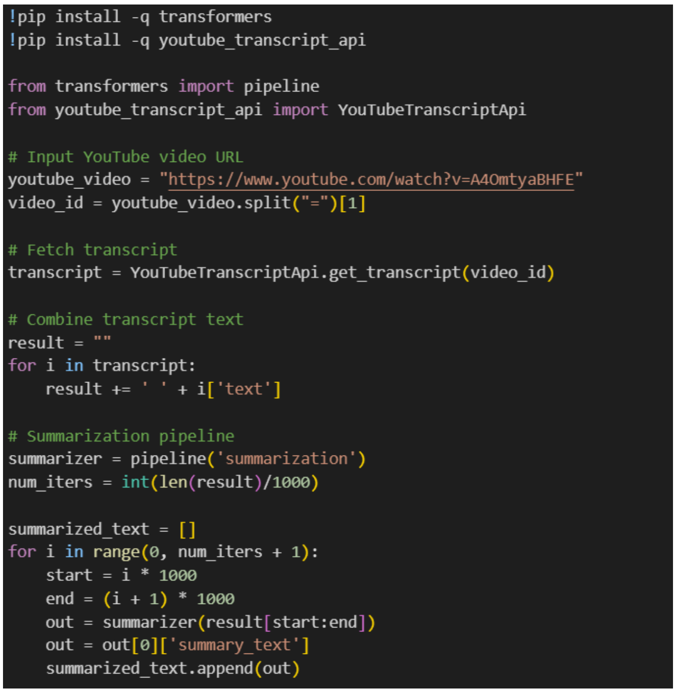
5)Assessment

The ultimate step includes assessing the quality of the created rundowns and interpretations. Usually accomplished utilizing two essential measurements. BLEU scores are calculated to measure the precision of interpretations by comparing them to reference writings. Moreover, human input is accumulated to assess coherence, semantic constancy, and lucidness. These assessments guarantee that the yield meets high-quality benchmarks and remains relevantly precise over dialects.

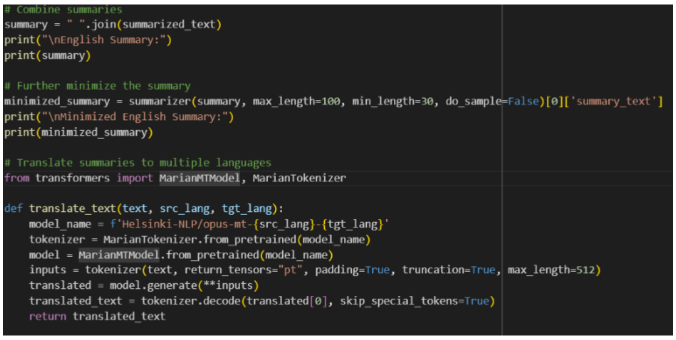
4. SOFTWARE REQUIREMENTS

OS: Windows, Language: Python 3.x, Libraries: (Hugging Face Transformers, Speech Recognition (for ASR), NLTK, Marian MT), Platform: Jupyter notebook, Google Collab.

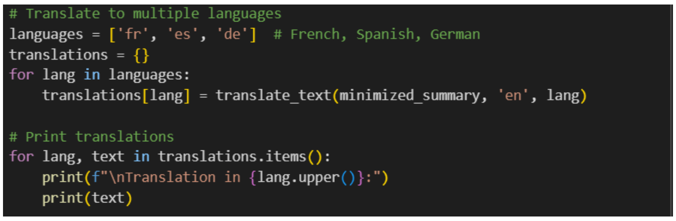
5. INPUT AND OUTPUT SCREENS



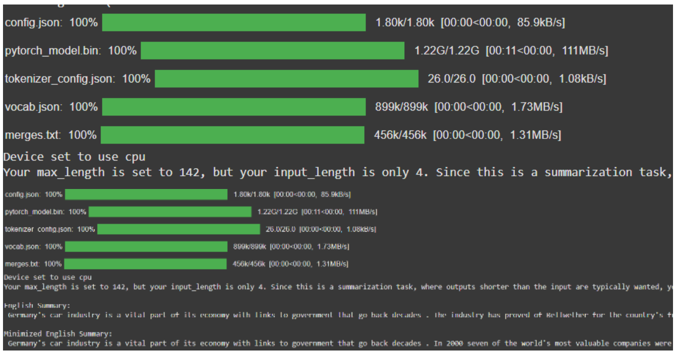
```
!pip install -q transformers
!pip install -q youtube_transcript_api

from transformers import pipeline
from youtube_transcript_api import YouTubeTranscriptApi

# Input YouTube video URL
youtube_video = "https://www.youtube.com/watch?v=A4OmtyaBHFE"
video_id = youtube_video.split("=")[1]

# Fetch transcript
transcript = YouTubeTranscriptApi.get_transcript(video_id)

# Combine transcript text
result = ""
for i in transcript:
    result += ' ' + i['text']

# Summarization pipeline
summarizer = pipeline('summarization')
num_iters = int(len(result)/1000)

summarized_text = []
for i in range(0, num_iters + 1):
    start = i * 1000
    end = (i + 1) * 1000
    out = summarizer(result[start:end])
    out = out[0]['summary_text']
    summarized_text.append(out)
```



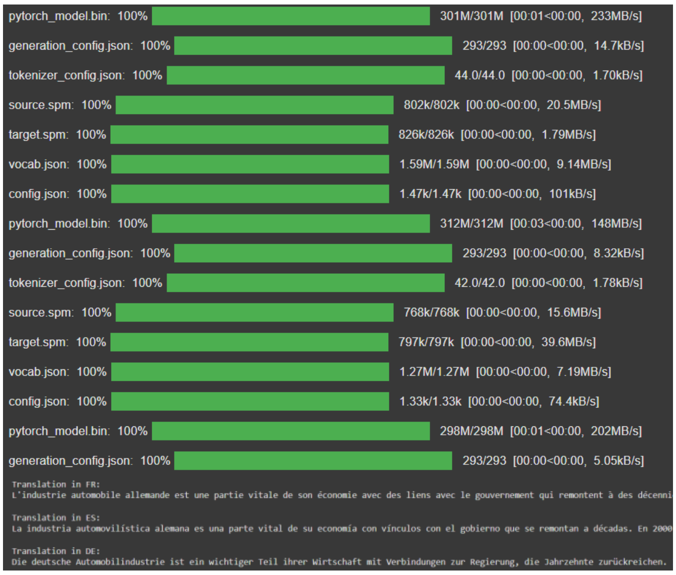
```
# Combine summaries
summary = " ".join(summarized_text)
print("\nEnglish Summary:")
print(summary)

# Further minimize the summary
minimized_summary = summarizer(summary, max_length=100, min_length=30, do_sample=False)[0]['summary_text']
print("\nMinimized English Summary:")
print(minimized_summary)

# Translate summaries to multiple languages
from transformers import MarianMTModel, MarianTokenizer

def translate_text(text, src_lang, tgt_lang):
    model_name = f'Helsinki-NLP/opus-mt-{src_lang}-{tgt_lang}'
    tokenizer = MarianTokenizer.from_pretrained(model_name)
    model = MarianMTModel.from_pretrained(model_name)
    inputs = tokenizer(text, return_tensors="pt", padding=True, truncation=True, max_length=512)
    translated = model.generate(**inputs)
    translated_text = tokenizer.decode(translated[0], skip_special_tokens=True)
    return translated_text
```



```
# Translate to multiple languages
languages = ['fr', 'es', 'de']  # French, Spanish, German
translations = {}
for lang in languages:
    translations[lang] = translate_text(minimized_summary, 'en', lang)

# Print translations
for lang, text in translations.items():
    print(f"\nTranslation in {lang.upper()}:")
    print(text)
```



```
config.json:  100%  1.80k/1.80k  [00:00<00:00,  85.9kB/s]
pytorch_model.bin:  100%  1.22G/1.22G  [00:11<00:00,  111MB/s]
tokenizer_config.json:  100%  26.0/26.0  [00:00<00:00,  1.08kB/s]
vocab.json:  100%  899k/899k  [00:00<00:00,  1.73MB/s]
merges.txt:  100%  456k/456k  [00:00<00:00,  1.31MB/s]
Device set to use cpu
Your max_length is set to 142, but your input_length is only 4. Since this is a summarization task,
config.json:  100%  1.80k/1.80k  [00:00<00:00,  85.9kB/s]
pytorch_model.bin:  100%  1.22G/1.22G  [00:11<00:00,  111MB/s]
tokenizer_config.json:  100%  26.0/26.0  [00:00<00:00,  1.08kB/s]
vocab.json:  100%  899k/899k  [00:00<00:00,  1.73MB/s]
merges.txt:  100%  456k/456k  [00:00<00:00,  1.31MB/s]
Device set to use cpu
Your max_length is set to 142, but your input_length is only 4. Since this is a summarization task, where outputs shorter than the input are typically wanted, y

English Summary:
 Germany's car industry is a vital part of its economy with links to government that go back decades . The industry has proved of Bellwether for the country's f

Minimized English Summary:
 Germany's car industry is a vital part of its economy with links to government that go back decades . In 2000 seven of the world's most valuable companies were
```



```
pytorch_model.bin:  100%  301M/301M  [00:01<00:00,  233MB/s]
generation_config.json:  100%  293/293  [00:00<00:00,  14.7kB/s]
tokenizer_config.json:  100%  44.0/44.0  [00:00<00:00,  1.70kB/s]
source.spm:  100%  802k/802k  [00:00<00:00,  20.5MB/s]
target.spm:  100%  826k/826k  [00:00<00:00,  1.79MB/s]
vocab.json:  100%  1.59M/1.59M  [00:00<00:00,  9.14MB/s]
config.json:  100%  1.47k/1.47k  [00:00<00:00,  101kB/s]
pytorch_model.bin:  100%  312M/312M  [00:03<00:00,  148MB/s]
generation_config.json:  100%  293/293  [00:00<00:00,  8.32kB/s]
tokenizer_config.json:  100%  42.0/42.0  [00:00<00:00,  1.78kB/s]
source.spm:  100%  768k/768k  [00:00<00:00,  15.6MB/s]
target.spm:  100%  797k/797k  [00:00<00:00,  39.6MB/s]
vocab.json:  100%  1.27M/1.27M  [00:00<00:00,  7.19MB/s]
config.json:  100%  1.33k/1.33k  [00:00<00:00,  74.4kB/s]
pytorch_model.bin:  100%  298M/298M  [00:01<00:00,  202MB/s]
generation_config.json:  100%  293/293  [00:00<00:00,  5.05kB/s]
 Translation in FR:
 L'industrie automobile allemande est une partie vitale de son économie avec des liens avec le gouvernement qui remontent à des décenni

 Translation in ES:
 La industria automovilística alemana es una parte vital de su economía con vínculos con el gobierno que se remontan a décadas. En 2000

 Translation in DE:
 Die deutsche Automobilindustrie ist ein wichtiger Teil ihrer Wirtschaft mit Verbindungen zur Regierung, die Jahrzehnte zurückreichen.
```

6. CONCLUSION

The framework viably extricates the transcript of a YouTube video, summarizes the substance, and deciphers the rundown into numerous dialects. Utilizing the YouTube Transcript API, it changes over the video's discourse into content. This content is at that point summarized into a brief adaptation utilizing the Embracing Confront summarization pipeline, holding basic subtle elements whereas decreasing repetition. The outline is advance minimized to guarantee brevity without losing basic data. Also, the minimized outline is deciphered into French, Spanish, and German utilizing Marian MT interpretation models, guaranteeing exact conservation of the center message. This approach streamlines the method of making video substance more available to a worldwide group of onlookers, and the system can be effectively amplified to bolster more dialects or particular utilize cases.

REFERENCES

1. Kadam, Payal, et al. "Recent Challenges and Opportunities in Video Summarization with Machine Learning Algorithms." IEEE Access 10 (2022): 122762-122785.

2. Singh, Yogendra, et al. "YouTube Video Summarizer using NLP: A Review." International Journal of Performability Engineering 19.12 (2023): 817.

3. Phani, Siginamsetty, et al. "MMSFT: Multilingual Multimodal Summarization by Fine-tuning Transformers." IEEE Access (2024).

4. Hande, Kapil, et al. "NLP based Video Summarisation using Machine Learning." (2023).

5. Gupta, Pooja, Swati Nigam, and Rajiv Singh. "Automatic Extractive Text Summarization using Multiple Linguistic Features." ACM Transactions on Asian and Low-Resource Language Information Processing (2024).

6. Li, Jinpeng, et al. "Multilingual Generation in Abstractive Summarization: A Comparative Study." Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024.