ABSTRACT

The amount of data available in a video is unpredictable hence video transcript summarization has become essential. Without video transcript summarization the ability to quickly grasp the information from a lengthy video becomes very inconvenient. YouTube has become an active platform in our life to seek information. Manually performing this task is very tiresome and time consuming task. Hence generating summaries for such lengthy videos will make the process convenient. The project consists of a transcript generator algorithm and a text summarizer followed by a language detector and translator. The entire simulation is carried out in an online website which allows us to generate a summary for any chosen YouTube video in real-time.

KEYWORDS: Natural Language Processing, Summarization, Language Conversion, Transformers.

INTRODUCTION

Since the invention of YouTube there has been an issue in discovering the important patterns of a video. The aim of YouTube Transcript summarization is to create a concise summary from a long video such that information is easier to grasp helping us save our efforts and time.

Need for Text Summarization

Summarization [1] would eliminate manual efforts. Shorter texts which are generated by summarizing larger texts would reduce reading time. With the ever-growing amount of data, text summarization would reduce the size which solves the problem of storage. A shorter summary provides more significant insights.

Summarization Techniques

There are two different approaches that are used for text summarization [2]

Extractive Summarization: This is where the model identifies the important sentences and phrases from the original text and only outputs those. [3]

Abstractive Summarization: The model produces a completely different text that is shorter than the original text, it generates new sentences in a new structure, like humans do. [4]

GoogleTransAPI

API stands for Application Programming Interface. It acts as an intermediate between two applications or software.

GoogleTrans is a free and unlimited python library that implements Google Translate API. This uses the Google Translate Ajax API to make calls to such methods like detect and translate.

Features:

× Fast and reliable – uses the same servers that translate.google.com does

× Auto Language Detection

× Bulk Translations

× HTTP/2 support

YouTube Transcript API

This is a python API which allows to get the transcript for any given YouTube video. It also works for automatically generated subtitles as well.

Getting list of all transcripts

× transcript.video\_id returns the ID of the video

× transcript.language returns us the language of the transcript

× transcript.langauge\_code returns the language code of the transcript

× transcript.is\_generated tells whether it has been manually created or generated by YouTube

SPACY-TRANSFORMER

PROBLEM STATEMENT

Nowadays students and just about any individual spends a considerable amount of their time on a YouTube video to gain an understanding of their curiosity. But there is a high probability that the video they have chosen doesn’t convey the information they are seeking or sometimes even be totally fraud. Which results in a total waste of their precious time and efforts. A repeated example of this can be, a student who wants to know whether the field of data science or computer science is more suitable for him. There could be many videos present on YouTube which are trying to analyze the differences and advantages of each field. But it is possible that all the information presented is already known by the viewer and doesn’t gain anything relevant from the video.

LITERATURE SURVEY

PROPOSED ARCHITECTURE

We utilized few pre-defined libraries to create a video transcript summarizer as an online website. The libraries used are YouTube transcript API, Transformers and GoogleTransAPI which are used to generate the transcript of any chosen YouTube video using the YouTube transcript API and then summarize the generated transcript using the Transformers algorithm. With this proposed system we generate a completely new summarized text that is different from the original text. Obtaining summary of lengthy videos to help quick grasp of information and helps us save time and efforts to go through the whole content. Language detection and conversion which makes it accessible to wide range of users.

YouTube transcript summarization is important where we want a quick grasp of the information a video of certain length holds without having the need to go through the whole duration of the video. Especially when we can translate the summary of the video into any language, we can grasp the information more effectively. There are no well-designed approaches to achieve this at a unitary place without complexities. The only inefficient method of gaining this service is by generating the transcript of a video and then summarizing it to later convert the language by using a language convertor.

CONCLUSION AND FUTURE WORK

We presented a secure YouTube transcript summarizer as an online website to simplify the process of checking the contents of a video and with the comfort of a language translator. In this paper, we

REFERENCES

1. Babar S, Tech-Cse M, RIT (2013) Text Summarization: An Overview

2. Allahyari M, Pouriyeh SA, Assefi M, Safaei S, Trippe ED, Gutierrez JB, Kochut K (2017) Text Summarization Techniques: A Brief Survey. CoRR abs/1707.02268:

3. Moratanch N, Gopalan C (2017) A survey on extractive text summarization. pp 1–6

4. Moratanch N, Gopalan C (2016) A survey on abstractive text summarization. pp 1–7