International Journal of Scientific and Research Publications, Volume X, Issue X, Month 2018 1 ISSN 2250-3153

**YouTube Transcript Summarization using Flask and T5 NLP Model**

**First Author\*, Second Author\*\*, Third Author\*\***

\* Department, Institute Name

\*\* Department, Institute Name, if any

***Abstract*-**YouTube Transcript Summarization is web application that is developed using Flask web framework and T5 model for summarizing the video transcript or subtitles. This application is useful for the student who wants to make notes of YouTube lectures as well as for all the people who uses YouTube for learning purpose

***Keywords***- YouTube, Transcript, Summarization, NLP, Flask, python, transformer, torch, REST, API.

I. INTRODUCTION

T

here are number of video lectures and recording being created and uploaded on internet throughout the day. Sometimes it is difficult to cover the whole video because we don’t have much time to view the whole video. Although if we can watch the whole video we couldn’t find relevant information out of it. So summarization of transcript or subtitles can solve this problem. YouTube Transcript Summarization is web application made using flask and T5 model that summarize the YouTube video transcript or subtitle. This application uses three dependencies flask , youtube\_transcript\_api and transformers[torch].Flask is micro web framework written in python which is used for making RESTful API apps. YTS is also made using flask web framework

----youtube transci api---- transformer

II. Literature survey

A literature review revealed research and studies based on the implementation of a summarization of text using Natural Language Processing and creating web application using Flask and also the different python dependencies required for making this web application i.e. YoutubeTranscriptApi ,Transformer , Flask, T5 pretrained model.

[1] Automated Text Summarization in SUMMARIST by Eduard Hovy and Chin-Yew Lin: They describe the about the Text Summarization equation i.e. summarization = topic identification + interpretation + generation and also about Text Pre-processing.

[2] Text Summarization ISSN: 2278-0181: This discuss about Text Summarization using Restricted Boltzmann Machine(RBM),Deep Neural Networks(DNN). [3] Text Summarization e-ISSN: 2395-0056 : They describe about summarization using Recurrent Neural Network(RNN). [3] Python Flask Framework by Vasanth

Nagarajan - A Step by Step Guide For the Beginners for learning flask web framework. [4] T5 by Hugging Face Community: Pretrained T5 model documentation. [5] Youtube-transcript-api: pypi.org.

III. System Objective

This application is used to summarize the transcript of youtube video by NLP techniques. The objective of this application is to help students to revise the lecture without actually seeing the whole video. Summarization of text is essential to get the important information while dealing with large collection of documents. With the advent of World Wide Web information has become intrinsic part of our life. To remember the details of every information is not possible for human mind. Therefore summarization of text documents plays a very important role in information gathering. Also the objective to make application

* Responsive
* User Friendly
* Robust

# APPLICATION ARCHITECTURE

YouTube Transcript Summarization application involves following step to perform summarization

1. **Render the home page**

The application first render the home page which have input field where user can type the youtube video url

@app.route('/home')

def home\_page():

return render\_template('home.html')

1. **Validating the URL**

In this step we will validate the URL whete

1. **Requesting Argument**

The application request for youtube video which is passed as argument in the URL.

args = request.args

youtube\_url =args.get("youtube\_url")

1. **Slice the videoID**

In this step the application slice or trim the url so that we can get videoID by calling the getVideoID() function. VideoID is a unique key used for identifying a particular video.

def getVideoId(youtube\_url):

id = youtube\_url.split("v=")

videoID = id[1][:11]

# IMPROVEMENT AS PER REVIEWER COMMENTS

# Analyze and understand all the provided review comments thoroughly. Now make the required amendments in your paper. If you are not confident about any review comment, then don't forget to get clarity about that comment. And in some cases there could be chances where your paper receives number of critical remarks. In that cases don't get disheartened and try to improvise the maximum.

# After submission IJSRP will send you reviewer comment within 10-15 days of submission and you can send us the updated paper within a week for publishing.

# This completes the entire process required for widespread of research work on open front. Generally all International Journals are governed by an Intellectual body and they select the most suitable paper for publishing after a thorough analysis of submitted paper. Selected paper get published (online and printed) in their periodicals and get indexed by number of sources. After the successful review and payment, IJSRP will publish your paper for the current edition. You can find the payment details at: <http://ijsrp.org/online-publication-charge.html>.

# CONCLUSION

# A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

# APPENDIX

# Appendixes, if needed, appear before the acknowledgment. ACKNOWLEDGMENT

# The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments.

# REFERENCES

# [1] G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

# [2] W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

# [3] H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.

# [4] B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.

# [5] E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” IEEE Trans. Antennas Propagat., to be published.

# [6] J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” IEEE J. Quantum Electron., submitted for publication. AUTHORS First Author – Author name, qualifications, associated institute (if any) and email address. Second Author – Author name, qualifications, associated institute (if any) and email address. Third Author – Author name, qualifications, associated institute (if any) and email address. Correspondence Author – Author name, email address, alternate email address (if any), contact number.