

Convert YouTube Videos into Subtitled Text with Whisper AI and Pytube



Anvi Kohli · [Follow](#)

3 min read · Aug 31, 2023



3



...

In this article, we'll dive into a project to convert YouTube videos into neatly subtitled text using the OpenAI's Whisper AI and Pytube. Whether you're a content enthusiast, a curious learner, or someone who loves to have subtitles on videos, this project empowers you to effortlessly transform spoken words into insightful, timestamped text. Let's Go!

Project Overview

The project involves the following steps:

1. Installing Dependencies: We'll install the Whisper AI library using the `git` command and the `pip` command for Pytube library for downloading YouTube videos.
2. Converting YouTube Video to Audio: Using Pytube, we'll download a YouTube video and convert it into an audio file (`.mp3` format).
3. Transcribing Audio: We'll use the Whisper AI tool to transcribe the audio file, generating a list of segments with timestamps and corresponding text.
4. Formatting and Saving: We'll format the transcribed segments with timestamps in the `[hh:mm:ss.sss --> hh:mm:ss.sss]` format and save the result in a text file.

Technologies and Libraries Used

- Whisper AI: Whisper AI is a state-of-the-art automatic speech recognition (ASR) system developed by OpenAI. It's designed to convert spoken language into written text.
- Pytube: The Pytube library provides a simple and efficient way to interact with YouTube videos, enabling us to download content for

further processing.

- Torch: Torch is used for neural network computations and forms an integral part of the Whisper AI framework.

Step-by-Step

Step 1: Installing Dependencies

Before we proceed, we need to install the necessary libraries:

```
!pip -qqq install git+https://github.com/openai/whisper.git  
!pip -qqq install pytube
```

Step 2: Importing Libraries and Loading the Model

```
from pytube import YouTube  
import whisper
```

```
import torch
import os

device = "cuda" if torch.cuda.is_available() else "cpu"
whisper_model = whisper.load_model("large", device=device)
```

Step 3: Converting YouTube Video to Audio

In this step, we'll use the Pytube library to download the YouTube video and convert it to audio:

```
def video_to_audio(video_URL, destination, final_filename):

    # Get the video
    video = YouTube(video_URL)

    # Convert video to Audio
    audio = video.streams.filter(only_audio=True).first()

    # Save to destination
    output = audio.download(output_path = destination)

    _, ext = os.path.splitext(output)
    new_file = final_filename + '.mp3'

    # Change the name of the file
    os.rename(output, new_file)
```

```
def convert(url):
    # Video to audio
    video_URL = url
    destination = "."
    final_filename = "audio_file_to_convert"
    video_to_audio(video_URL, destination, final_filename)
```

Step 4: Transcribing the Audio

We'll leverage Whisper AI to transcribe the audio from the downloaded video:

```
def transcribe():
    audio_file = "audio_file_to_convert.mp3"
    result = whisper_model.transcribe(audio_file)
    result_segments = result['segments']
    print(result_segments)
    return format_segments(result_segments)
```

Step 5: Formatting and Saving

We'll format the transcribed segments and save them to a text file:

```
def format_segments(result_segments):
    formatted_output = []

    for segment in result_segments:
        start_time = segment['start']
        end_time = segment['end']
        text = segment['text']

        formatted_text = f"[{format_time_milliseconds(start_time)} --> {format_time_milliseconds(end_time)}] {text}"
        formatted_output.append(formatted_text)

    return "\n".join(formatted_output)

def format_time_milliseconds(seconds):
    minutes, seconds = divmod(seconds, 60)
    hours, minutes = divmod(minutes, 60)
    milliseconds = int((seconds - int(seconds)) * 1000)
    return f"{int(hours):01}:{int(minutes):01}:{int(seconds):02}.{milliseconds:03}
```

Function to save formatted text into a txt file

```
# Save the formatted result to a text file
def dump_into_txt(formatted_result):
    output_file_path = 'transcribed_text.txt'
    with open(output_file_path, 'w') as output_file:
        output_file.write(formatted_result)
    print(f"Formatted result saved to {output_file_path}")
```

And that's it. Here is the simple way you can transcribe any video from YouTube and generate its transcripts. Not just YouTube, but the WhisperAI can be used to transcribe any video in any format.

The screenshot shows a terminal window with two tabs: '+ Code' and '+ Text'. The '+ Code' tab contains the following Python code:

```
# Save the formatted result to a text file
def dump_into_txt(formatted_result):
    output_file_path = 'transcribed_text.txt'
    with open(output_file_path, 'w') as output_file:
        output_file.write(formatted_result)
    print(f"Formatted result saved to {output_file_path}")
```

The '+ Text' tab shows the execution of this code. It starts with the instruction 'Main Function to Run' followed by the definition of the `main()` function:

```
[9] def main():
    url = input("Enter Link of Youtube Video to be converted: ")
    convert(url)
    formatted = transcribe()
    dump_into_txt(formatted)
```

Then, the command `[10] main()` is run, which prompts for a YouTube URL and prints the transcribed text:

```
Enter Link of Youtube Video to be converted: https://www.youtube.com/watch?v=v394vC23Hw0
[{'id': 0, 'seek': 0, 'start': 0.0, 'end': 8.0, 'text': "It's 2015, and the world's most powerful drug trafficker, El Chapo, is about to escape from pri
Formatted result saved to transcribed_text.txt
```

The terminal also displays a progress bar and the completion message '3m 7s completed at 1:01 PM'.

Main Function

```
transcribed_text.txt x
1 --> 0:0:08.000] It's 2015, and the world's most powerful drug trafficker, El Chapo, is about to escape from prison.
2 --> 0:0:12.000] Did he see it?
3 --> 0:0:14.000] Right now.
4 --> 0:0:18.000] He's gone.
5 --> 0:0:29.000] One of the world's richest, most notorious drug kingpins broke out of prison.
6 --> 0:0:34.000] The former cartel leader is considered one of the most dangerous drug traffickers in the world.
7 --> 0:0:40.000] Joaquin Guzman Loera, better known as El Chapo, broke out of his cell by crawling through a hole in the shower.
8 --> 0:0:46.000] The escape of Chapo Guzman translates into two things.
9 --> 0:0:51.000] More violence for Mexico, and more drugs coming into the United States.
10 --> 0:0:59.000] El Chapo is the leader of the Sinaloa Cartel, a murderous criminal organization that has killed thousands during Mexico's violent drug war.
11 --> 0:1:03.000] Their methods are brutal.
12 --> 0:1:09.000] But violence isn't what makes El Chapo different from other traffickers.
13 --> 0:1:13.000] Instead, his legacy has been shaped by two abilities.
14 --> 0:1:17.000] Transporting more drugs across the border than anyone else.
15 --> 0:1:21.000] And always finding a way to escape authorities.
16 --> 0:1:24.000] That was, until now.
17 --> 0:1:29.000] After being captured again in Mexico, El Chapo is standing trial in the United States.
18 --> 0:1:39.000] It's the latest chapter for the world's most powerful drug trafficker, and to understand how El Chapo got here, it helps to look at a single tool.
19 --> 0:1:46.000] The one tool that El Chapo used to transform the drug trade is the same tool that made him its most elusive target.
20 --> 0:1:48.000] Tunnels. Tunnels. Tunnels. Through a tunnel.
21 --> 0:1:54.000] Used by El Chapo. Tunnel king. The master of tunnels. El rey de los tunneles. Prince of tunnels.
22 --> 0:1:58.000] The magnificent feat of architecture and engineering. Being used to smuggle drugs.
23 --> 0:2:00.000] And it's like looking for a needle in a haystack.
24 --> 0:2:04.000] There's not really any form of technology right now that can discover these tunnels being built.
25 --> 0:2:06.000] The tub. Look at this.
26 --> 0:2:14.000] In 1990, federal agents in Douglas, Arizona found something new.
27 --> 0:2:17.000] A sophisticated drug tunnel to the U.S. from Mexico.
28 --> 0:2:22.000] And when investigators got inside, they were amazed by what they saw.
29 --> 0:2:27.000] The tunnel was 300 feet long and lit by an electrical system running along the wall.
30 --> 0:2:33.000] At one end, a lever raised a hydraulic lift that opened an entrance in the safehouse floor.
31 --> 0:2:38.000] One agent said it was like something out of a James Bond movie.
32 --> 0:2:42.000] In fact, it had been built by an architect working for El Chapo
33 --> 0:2:42.000] ✓ 3m 7s completed at 1:01 PM
```

Generated Text File using the code

The code of the above article can be found here: [anvichip/yt-video-audio2text-whisper \(github.com\)](#)

Happy Learning!

Medium

Search

Write



Whisper Ai

YouTube

Pytube

Python

OpenAI



Written by Anvi Kohli

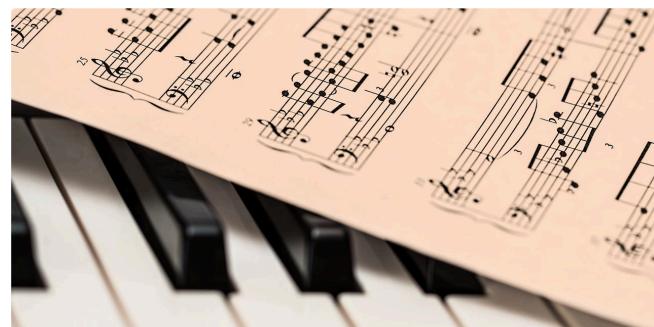
2 Followers

Techie | Budding AI/ML scientist | Healthcare

Follow



More from Anvi Kohli



The thumbnail features a pink vertical bar on the left and a white bar on the right. At the top is a black smiley face icon. Below it, the word "TECHNOLOGY" is written in bold capital letters. A horizontal line separates this from the title "Architecting", which is also in bold capital letters.


 Anvi Kohli

Get Images & Videos for free from keywords using Pixabay API

Visual content plays a pivotal role in captivating audiences and conveying...

Mar 26



...


 Anvi Kohli

Navigating the Cloud: A Journey through AWS Solutions Architect...

Empowering Architects to Design Resilient Cloud Solutions

Dec 17, 2023



1



...



 Anvi Kohli

Generate Images using Stable Diffusion XL model

Welcome to the Part—1 of the Stable Diffusion Series where we today we will chec...

May 2



1



...


 Anvi Kohli

How to Load Kaggle Datasets directly onto Google Colab—3...

Tired of downloading large datasets locally to run large AI models? Well, thankfully for us ...

Mar 30



...

See all from Anvi Kohli

Recommended from Medium



 Mark O'Brien

Whisper & Python for Video Transcription

Explore the Whisper Model from OpenAI for audio transcription made easy, fully local

 Mar 4  6

  ...

 Jeremy Savage

Using OpenAI's Whisper to Transcribe Real-time Audio

The availability of advanced technology and tools, in particular, AI is increasing at an ever...

Apr 12  72

  ...

Lists



Coding & Development

11 stories · 694 saves



Predictive Modeling w/ Python

20 stories · 1364 saves



Practical Guides to Machine Learning

10 stories · 1645 saves



ChatGPT

21 stories · 711 saves



Dean Martin

Text to Speech from C# using and XTTs v2 (Python), with Chains &...

In this post we build on the previous post to build a completely free and state-of-the-art...

Jan 22 21

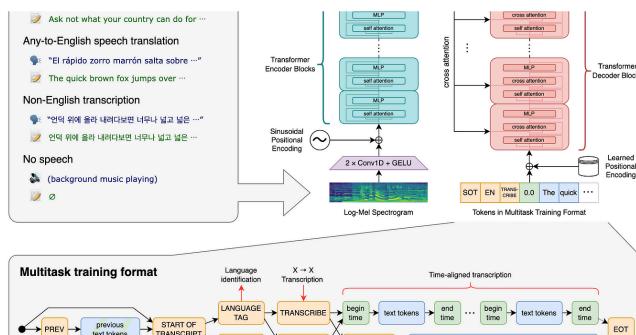


Imane Momayiz

Scraping Youtube video transcripts

How to fetch video transcripts from any YouTube channel with YouTube Data API and...

Mar 14



Benjamin Consolvo in Intel Analytics Software

Automatic Speech Recognition using OpenAI Whisper without a...



Be Tech! with Santander in Be Tech! with Santander

Guide: The Rise of Voice Cloning Technology

Easy Step-by-Step Guide to English and French Transcription and Translation on CPUs

Mar 14 · 211 views · 4 comments



...

By Vicentes Motos. Discover how to generate realistically human voices using a variety of...

Feb 22 · 57 views



...

See more recommendations

[Help](#) [Status](#) [About](#) [Careers](#) [Press](#) [Blog](#) [Privacy](#) [Terms](#) [Text to speech](#) [Teams](#)