YouTubeTranscriptApi

<https://www.youtube.com/watch?v=8ghTWrz-rrA>

## **Project description**

# **YouTube Transcript/Subtitle API (including automatically generated subtitles and subtitle translations)**

**Donate** Build Status Coverage Status MIT license image image

This is a python API which allows you to get the transcripts/subtitles for a given YouTube video. It also works for automatically generated subtitles, supports translating subtitles and it does not require a headless browser, like other selenium based solutions do!

## **Install**

It is recommended to [install this module by using pip](https://pypi.org/project/youtube-transcript-api/):

pip install youtube\_transcript\_api

You can either integrate this module [into an existing application](https://pypi.org/project/youtube-transcript-api/#api), or just use it via an [CLI](https://pypi.org/project/youtube-transcript-api/#cli).

## **API**

The easiest way to get a transcript for a given video is to execute:

from youtube\_transcript\_api import YouTubeTranscriptApi

YouTubeTranscriptApi.get\_transcript(video\_id)

3Rzi11Hvyh0

This will return a list of dictionaries looking somewhat like this:

[

{

'text': 'Hey there',

'start': 7.58,

'duration': 6.13

},

{

'text': 'how are you',

'start': 14.08,

'duration': 7.58

},

# ...

]

You can also add the languages param if you want to make sure the transcripts are retrieved in your desired language (it defaults to english).

YouTubeTranscriptApi.get\_transcripts(video\_ids, languages=['de', 'en'])

It's a list of language codes in a descending priority. In this example it will first try to fetch the german transcript ('de') and then fetch the english transcript ('en') if it fails to do so. If you want to find out which languages are available first, [have a look at list\_transcripts()](https://pypi.org/project/youtube-transcript-api/#list-available-transcripts)

To get transcripts for a list of video ids you can call:

YouTubeTranscriptApi.get\_transcripts(video\_ids, languages=['de', 'en'])

### **By example**

# retrieve the available transcripts

transcript\_list = YouTubeTranscriptApi.list\_transcripts('video\_id')

# iterate over all available transcripts

for transcript in transcript\_list:

# the Transcript object provides metadata properties

print(

transcript.video\_id,

transcript.language,

transcript.language\_code,

# whether it has been manually created or generated by YouTube

transcript.is\_generated,

# whether this transcript can be translated or not

transcript.is\_translatable,

# a list of languages the transcript can be translated to

transcript.translation\_languages,

)

# fetch the actual transcript data

print(transcript.fetch())

# translating the transcript will return another transcript object

print(transcript.translate('en').fetch())

# you can also directly filter for the language you are looking for, using the transcript list

transcript = transcript\_list.find\_transcript(['de', 'en'])

# or just filter for manually created transcripts

transcript = transcript\_list.find\_manually\_created\_transcript(['de', 'en'])

# or automatically generated ones

transcript = transcript\_list.find\_generated\_transcript(['de', 'en'])