pytube Documentation

Release 10.7.1

Nick Ficano

Contents

1	Featu	ires	3
2 The User Guide			
	2.1	Installation of pytube	5
	2.2		6
	2.3		6
	2.4	Filtering Streams	7
	2.5		8
	2.6	Using Playlists	9
	2.7	Command-line interface (CLI)	10
	2.8	Exception handling	
3	The A	API Documentation	13
	3.1	API	13
4	Indic	es and tables	29
Ру	thon N	Module Index	31
			22
In	dex		33

Release v10.7.1. (Installation)

pytube is a lightweight, Pythonic, dependency-free, library (and command-line utility) for downloading YouTube Videos.

Behold, a perfect balance of simplicity versus flexibility:

```
>>> from pytube import YouTube
>>> YouTube('https://youtu.be/9bZkp7q19f0').streams.first().download()
>>> yt = YouTube('http://youtube.com/watch?v=9bZkp7q19f0')
>>> yt.streams
... .filter(progressive=True, file_extension='mp4')
... .order_by('resolution')
... .desc()
... .first()
... .download()
```

Contents 1

2 Contents

CHAPTER 1

Features

- Support for Both Progressive & DASH Streams
- $\bullet \ \ Easily \ Register \ \texttt{on_download_progress} \ \& \ \texttt{on_download_complete} \ callbacks$
- Command-line Interfaced Included
- Caption Track Support
- Outputs Caption Tracks to .srt format (SubRip Subtitle)
- Ability to Capture Thumbnail URL.
- Extensively Documented Source Code
- No Third-Party Dependencies

4 Chapter 1. Features

CHAPTER 2

The User Guide

This part of the documentation begins with some background information about the project, then focuses on step-by-step instructions for getting the most out of pytube.

2.1 Installation of pytube

This guide assumes you already have python and pip installed.

To install pytube, run the following command in your terminal:

```
$ pip install pytube
```

2.1.1 Get the Source Code

pytube is actively developed on GitHub, where the source is available.

You can either clone the public repository:

```
$ git clone git://github.com/pytube/pytube.git
```

Or, download the tarball:

```
$ curl -OL https://github.com/pytube/pytube/tarball/master # optionally, zipball is also available (for Windows users).
```

Once you have a copy of the source, you can embed it in your Python package, or install it into your site-packages by running:

```
$ cd pytube
$ python -m pip install .
```

2.2 Quickstart

This guide will walk you through the basic usage of pytube.

Let's get started with some examples.

2.2.1 Downloading a Video

Downloading a video from YouTube with pytube is incredibly easy.

Begin by importing the YouTube class:

```
>>> from pytube import YouTube
```

Now, let's try to download a video. For this example, let's take something like the YouTube Rewind video for 2019:

```
>>> yt = YouTube('http://youtube.com/watch?v=21Ae1cqCOXo')
```

Now, we have a YouTube object called yt.

The pytube API makes all information intuitive to access. For example, this is how you would get the video's title:

```
>>> yt.title
YouTube Rewind 2019: For the Record | #YouTubeRewind
```

And this would be how you would get the thumbnail url:

Neat, right? For advanced use cases, you can provide some additional arguments when you create a YouTube object:

```
>>> yt = YouTube(
    'http://youtube.com/watch?v=21Ae1cqCOXo',
    on_progress_callback=progress_func,
    on_complete_callback=complete_func,
    proxies=my_proxies
)
```

When instantiating a YouTube object, these named arguments can be passed in to improve functionality.

The on_progress_callback function will run whenever a chunk is downloaded from a video, and is called with three arguments: the stream, the data chunk, and the bytes remaining in the video. This could be used, for example, to display a progress bar.

The on_complete_callback function will run after a video has been fully downloaded, and is called with two arguments: the stream and the file path. This could be used, for example, to perform post-download processing on a video like trimming the length of it.

Once you have a YouTube object set up, you're ready to start looking at different media streams for the video, which is discussed in the next section.

2.3 Working with Streams and StreamQuery

The next section will explore the various options available for working with media streams, but before we can dive in, we need to review a new-ish streaming technique adopted by YouTube. It assumes that you have already created a YouTube object in your code called "yt".

2.3.1 DASH vs Progressive Streams

Begin by running the following to list all streams:

You may notice that some streams listed have both a video codec and audio codec, while others have just video or just audio, this is a result of YouTube supporting a streaming technique called Dynamic Adaptive Streaming over HTTP (DASH).

In the context of pytube, the implications are for the highest quality streams; you now need to download both the audio and video tracks and then post-process them with software like FFmpeg to merge them.

The legacy streams that contain the audio and video in a single file (referred to as "progressive download") are still available, but only for resolutions 720p and below.

2.4 Filtering Streams

Pytube has built-in functionality to filter the streams available in a YouTube object with the .filter() method. You can pass it a number of different keyword arguments, so let's review some of the different options you're most likely to use. For a complete list of available properties to filter on, you can view the API documentation here: pytube. StreamQuery.filter().

2.4.1 Filtering by streaming method

As mentioned before, progressive streams have the video and audio in a single file, but typically do not provide the highest quality media; meanwhile, adaptive streams split the video and audio tracks but can provide much higher quality. Pytube makes it easy to filter based on the type of stream that you're interested.

For example, you can filter to only progressive streams with the following:

Conversely, if you only want to see the DASH streams (also referred to as "adaptive") you can do:

```
>>> yt.streams.filter(adaptive=True)

[<Stream: itag="137" mime_type="video/mp4" res="1080p" fps="30fps" vcodec="avc1.640028

... progressive="False" type="video">,

<Stream: itag="248" mime_type="video/webm" res="1080p" fps="30fps" vcodec="vp9"

... progressive="False" type="video">,

<Stream: itag="399" mime_type="video/mp4" res="None" fps="30fps" vcodec="av01.0.08M.08

... progressive="False" type="video">,

...

<Stream: itag="250" mime_type="audio/webm" abr="70kbps" acodec="opus" progressive=

... "False" type="audio">,

<Stream: itag="251" mime_type="audio/webm" abr="160kbps" acodec="opus" progressive=

... "False" type="audio">]
```

2.4.2 Filtering for audio-only streams

To query the streams that contain only the audio track:

2.4.3 Filtering for MP4 streams

To query only streams in the MP4 format:

2.5 Subtitle/Caption Tracks

Pytube exposes the caption tracks in much the same way as querying the media streams. Let's begin by switching to a video that contains them:

```
>>> yt = YouTube('http://youtube.com/watch?v=21Ae1cqCOXo')
>>> yt.captions
{'ar': <Caption lang="Arabic" code="ar">, 'zh-HK': <Caption lang="Chinese (Hong Kong)
\hookrightarrow" code="zh-HK">, 'zh-TW': <Caption lang="Chinese (Taiwan)" code="zh-TW">, 'hr':
→ <Caption lang="Croatian" code="hr">, 'cs': <Caption lang="Czech" code="cs">, 'da':
→ < Caption lang="Danish" code="da">, 'nl': < Caption lang="Dutch" code="nl">, 'en':
→ < Caption lang="English" code="en">, 'en-GB': < Caption lang="English" (United Kingdom)
→" code="en-GB">, 'et': <Caption lang="Estonian" code="et">, 'fil': <Caption lang=
→ "Filipino" code="fil">, 'fi': <Caption lang="Finnish" code="fi">, 'fr-CA': <Caption...
→lang="French (Canada)" code="fr-CA">, 'fr-FR': <Caption lang="French (France)" code=
→"fr-FR">, 'de': <Caption lang="German" code="de">, 'el': <Caption lang="Greek" code=
→"el">, 'iw': <Caption lang="Hebrew" code="iw">, 'hu': <Caption lang="Hungarian"...
→code="hu">, 'id': <Caption lang="Indonesian" code="id">, 'it': <Caption lang=
→"Italian" code="it">, 'ja': <Caption lang="Japanese" code="ja">, 'ko': <Caption_
→lang="Korean" code="ko">, 'lv': <Caption lang="Latvian" code="lv">, 'lt': <Caption...
→lang="Lithuanian" code="lt">, 'ms': <Caption lang="Malay" code="ms">, 'no':
→ <Caption lang="Norwegian" code="no">, 'pl': <Caption lang="Polish" code="pl">, 'pt-
→BR': <Caption lang="Portuguese (Brazil)" code="pt-BR">, 'pt-PT': <Caption lang=
→ "Portuguese (Portugal)" code="pt-PT">, 'ro': <Caption lang="Romanian" code="ro">,
→'ru': <Caption lang="Russian" code="ru">, 'sk': <Caption lang="Slovak" code="sk">,
→'es-419': <Caption lang="Spanish (Latin America)" code="es-419">, 'es-ES': <Caption,
→lang="Spanish (Spain)" code="es-ES">, 'sv': <Caption lang="Swedish" code="sv">, 'th
→': <Caption lang="Thai" code="th">, 'tr': <Caption lang="Turkish" code="tr">, 'uk':
→ <Caption lang="Ukrainian" code="uk">, 'ur': <Caption lang="Urdu" code="ur">, 'vi':
→ < Caption lang="Vietnamese" code="vi">}
```

Now let's checkout the english captions:

```
>>> caption = yt.captions.get_by_language_code('en')
```

Great, now let's see how YouTube formats them:

Oh, this isn't very easy to work with, let's convert them to the srt format:

```
>>> print(caption.generate_srt_captions())
1
00:00:10,200 --> 00:00:11,140
K-pop!
2
00:00:13,400 --> 00:00:16,200
That is so awkward to watch.
...
```

2.6 Using Playlists

This guide will walk you through the basics of working with pytube Playlists.

2.6.1 Creating a Playlist

Using pytube to interact with playlists is very simple. Begin by importing the Playlist class:

```
>>> from pytube import Playlist
```

Now let's create a playlist object. You can do this by initializing the object with a playlist URL:

```
>>> p = Playlist('https://www.youtube.com/playlist?list=PLS1QulWo1RIaJECMeUT4LFwJ-

-ghgoSH6n')
```

Or you can create one from a video link in a playlist:

Now, we have a Playlist object called p that we can do some work with.

2.6.2 Interacting with a playlist

Fundamentally, a Playlist object is just a container for YouTube objects.

If, for example, we wanted to download all of the videos in a playlist, we would do the following:

```
>>> print(f'Downloading: {p.title}')
Downloading: Python Tutorial for Beginers (For Absolute Beginners)
>>> for video in p.videos:
>>> video.streams.first().download()
```

Or, if we're only interested in the URLs for the videos, we can look at those as well:

And that's basically all there is to it! The Playlist class is relatively straightforward.

2.7 Command-line interface (CLI)

Pytube also ships with a tiny CLI for interacting with videos and playlists.

To download the highest resolution progressive stream:

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo
```

To view available streams:

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo --list
```

To download a specific stream, use the itag

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo --itag=22
```

To get a list of all subtitles (caption codes)

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo --list-captions
```

To download a specific subtitle (caption code) - in this case the English subtitles (in srt format) - use:

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo -c en
```

It is also possible to just download the audio stream (default AAC/mp4):

```
$ pytube https://www.youtube.com/watch?v=21Ae1cqCOXo -a
```

To list all command line options, simply type

```
$ pytube --help
```

Finally, if you're filing a bug report, the cli contains a switch called --build-playback-report, which bundles up the state, allowing others to easily replay your issue.

2.8 Exception handling

Pytube implements a number of useful exceptions for handling program flow. There are a number of cases where pytube simply cannot access videos on YouTube and relies on the user to handle these exceptions. Generally speaking, if a video is unaccessible for any reason, this can be caught with the generic VideoUnavailable exception. This could be used, for example, to skip private videos in a playlist, videos that are region-restricted, and more.

Let's see what your code might look like if you need to do exception handling:

This will automatically skip over videos that could not be downloaded due to a limitation with the pytube library. You can find more details about what specific exceptions can be handled here: pytube.exceptions.

CHAPTER 3

The API Documentation

If you are looking for information on a specific function, class, or method, this part of the documentation is for you.

3.1 API

3.1.1 YouTube Object

```
class pytube.YouTube(url: str, on_progress_callback: Optional[Callable[[Any, bytes, int], None]] =
                            None, on_complete_callback: Optional[Callable[[Any, Optional[str]], None]]
                             = None, proxies: Dict[str, str] = None)
     Core developer interface for pytube.
     author
          Get the video author. :rtype: str
     caption_tracks
          Get a list of Caption.
               Return type List[Caption]
     captions
          Interface to query caption tracks.
               Return type CaptionQuery.
     check_availability()
          Check whether the video is available.
          Raises different exceptions based on why the video is unavailable, otherwise does nothing.
     description
          Get the video description.
               Return type str
```

fmt streams

Returns a list of streams if they have been initialized.

If the streams have not been initialized, finds all relevant streams and initializes them.

keywords

Get the video keywords. :rtype: List[str]

length

Get the video length in seconds.

Return type int

metadata

Get the metadata for the video.

Return type YouTubeMetadata

player_response

The player response contains subtitle information and video details.

publish_date

Get the publish date.

Return type datetime

rating

Get the video average rating.

Return type float

register_on_complete_callback (func: Callable[[Any, Optional[str]], None])

Register a download complete callback function post initialization.

Parameters func (callable) - A callback function that takes stream and file_path.

Return type None

register_on_progress_callback (func: Callable[[Any, bytes, int], None])

Register a download progress callback function post initialization.

Parameters func (callable) -

A callback function that takes stream, chunk, and bytes_remaining as parameters.

Return type None

streams

Interface to query both adaptive (DASH) and progressive streams.

Return type StreamQuery.

thumbnail_url

Get the thumbnail url image.

Return type str

title

Get the video title.

Return type str

vid_info

Parse the raw vid info and return the parsed result.

Return type Dict[Any, Any]

views

Get the number of the times the video has been viewed.

Return type int

3.1.2 Playlist Object

```
class pytube.contrib.playlist.Playlist (url: str, proxies: Optional[Dict[str, str]] = None)
    Load a YouTube playlist with URL
```

 $\verb"count" (value") \rightarrow integer-return number of occurrences of value$

index ($value[, start[, stop]]) \rightarrow integer - return first index of value.$

Raises ValueError if the value is not present.

Supporting start and stop arguments is optional, but recommended.

title

Extract playlist title

Returns playlist title (name)

Return type Optional[str]

trimmed(*video_id: str*) → Iterable[str]

Retrieve a list of YouTube video URLs trimmed at the given video ID

i.e. if the playlist has video IDs 1,2,3,4 calling trimmed(3) returns [1,2] :type video_id: str video ID to trim the returned list of playlist URLs at

Return type List[str]

Returns List of video URLs from the playlist trimmed at the given ID

video_urls

Complete links of all the videos in playlist

Return type List[str]

Returns List of video URLs

videos

Yields YouTube objects of videos in this playlist

Yields YouTube

3.1.3 Stream Object

```
class pytube.Stream(stream: Dict[KT, VT], player_config_args: Dict[KT, VT], monostate: pytube.monostate.Monostate)
```

Container for stream manifest data.

default filename

Generate filename based on the video title.

Return type str

Returns An os file system compatible filename.

download (output_path: Optional[str] = None, filename: Optional[str] = None, filename_prefix: Optional[str] = None, skip_existing: bool = True, timeout: Optional[int] = None, max_retries: Optional[int] = 0) \rightarrow str Write the media stream to disk.

Parameters

- output_path (str or None) (optional) Output path for writing media file. If one is not specified, defaults to the current working directory.
- **filename** (*str or None*) (optional) Output filename (stem only) for writing media file. If one is not specified, the default filename is used.
- **filename_prefix** (str or None) (optional) A string that will be prepended to the filename. For example a number in a playlist or the name of a series. If one is not specified, nothing will be prepended This is separate from filename so you can use the default filename but still add a prefix.
- skip_existing (bool) (optional) Skip existing files, defaults to True
- timeout (int) (optional) Request timeout length in seconds

Returns Path to the saved video

Return type str

filesize

File size of the media stream in bytes.

Return type int

Returns Filesize (in bytes) of the stream.

filesize_approx

Get approximate filesize of the video

Falls back to HTTP call if there is not sufficient information to approximate

Return type int

Returns size of video in bytes

includes_audio_track

Whether the stream only contains audio.

Return type bool

includes_video_track

Whether the stream only contains video.

Return type bool

is adaptive

Whether the stream is DASH.

Return type bool

is_progressive

Whether the stream is progressive.

Return type bool

on_complete (file_path: Optional[str])

On download complete handler function.

Parameters file_path (str) – The file handle where the media is being written to.

Return type None

```
on_progress (chunk: bytes, file_handler: BinaryIO, bytes_remaining: int)
On progress callback function.
```

This function writes the binary data to the file, then checks if an additional callback is defined in the monostate. This is exposed to allow things like displaying a progress bar.

Parameters

- **chunk** (bytes) Segment of media file binary data, not yet written to disk.
- file_handler (io.BufferedWriter) The file handle where the media is being written to.
- bytes_remaining (int) The delta between the total file size in bytes and amount already downloaded.

Return type None

```
\textbf{parse\_codecs} \ () \ \rightarrow Tuple[Optional[str], Optional[str]]
```

Get the video/audio codecs from list of codecs.

Parse a variable length sized list of codecs and returns a constant two element tuple, with the video codec as the first element and audio as the second. Returns None if one is not available (adaptive only).

Return type tuple

Returns A two element tuple with audio and video codecs.

```
stream\_to\_buffer(buffer: BinaryIO) \rightarrow None
```

Write the media stream to buffer

Return type io.BytesIO buffer

title

Get title of video

Return type str

Returns Youtube video title

3.1.4 StreamQuery Object

```
class pytube.query.StreamQuery(fmt_streams)
```

Interface for querying the available media streams.

all () \rightarrow List[pytube.streams.Stream]

Get all the results represented by this query as a list.

Return type list

 $\verb"asc"() \to pytube.query. Stream Query"$

Sort streams in ascending order.

Return type StreamQuery

count (value: Optional[str] = None) \rightarrow int

Get the count of items in the list.

Return type int

 $desc() \rightarrow pytube.query.StreamQuery$

Sort streams in descending order.

Return type StreamQuery

filter(fps=None, res=None, resolution=None, mime_type=None, type=None, subtype=None,
 file_extension=None, abr=None, bitrate=None, video_codec=None, audio_codec=None,
 only_audio=None, only_video=None, progressive=None, adaptive=None, is_dash=None, custom_filter_functions=None)

Apply the given filtering criterion.

Parameters

- **fps** (*int* or *None*) (optional) The frames per second.
- resolution (str or None) (optional) Alias to res.
- res (str or None) (optional) The video resolution.
- mime_type (str or None) (optional) Two-part identifier for file formats and format contents composed of a "type", a "subtype".
- **type** (*str or None*) (optional) Type part of the mime_type (e.g.: audio, video).
- **subtype** (*str or None*) (optional) Sub-type part of the mime_type (e.g.: mp4, mov).
- **file_extension** (str or None) (optional) Alias to sub_type.
- **abr** (str or None) (optional) Average bitrate (ABR) refers to the average amount of data transferred per unit of time (e.g.: 64kbps, 192kbps).
- bitrate (str or None) (optional) Alias to abr.
- video_codec (str or None) (optional) Video compression format.
- audio_codec (str or None) (optional) Audio compression format.
- **progressive** (bool) Excludes adaptive streams (one file contains both audio and video tracks).
- **adaptive** (bool) Excludes progressive streams (audio and video are on separate tracks).
- is_dash (bool) Include/exclude dash streams.
- only_audio (bool) Excludes streams with video tracks.
- only_video (bool) Excludes streams with audio tracks.
- **custom_filter_functions** (*list or None*) (optional) Interface for defining complex filters without subclassing.

$first() \rightarrow Optional[pytube.streams.Stream]$

Get the first Stream in the results.

Return type Stream or None

Returns the first result of this query or None if the result doesn't contain any streams.

get_audio_only (subtype: str = 'mp4') \rightarrow Optional[pytube.streams.Stream]

Get highest bitrate audio stream for given codec (defaults to mp4)

Parameters subtype (str) – Audio subtype, defaults to mp4

Return type Stream or None

Returns The Stream matching the given itag or None if not found.

 $\texttt{get_by_itag}(itag: int) \rightarrow \text{Optional[pytube.streams.Stream]}$ Get the corresponding Stream for a given itag.

```
Return type Stream or None
               Returns The Stream matching the given itag or None if not found.
     get_by_resolution (resolution: str) → Optional[pytube.streams.Stream]
           Get the corresponding Stream for a given resolution.
           Stream must be a progressive mp4.
               Parameters resolution (str) - Video resolution i.e. "720p", "480p", "360p", "240p",
                   "144p"
               Return type Stream or None
               Returns The Stream matching the given itag or None if not found.
     \verb"get_highest_resolution"() \rightarrow Optional[pytube.streams.Stream]
           Get highest resolution stream that is a progressive video.
               Return type Stream or None
               Returns The Stream matching the given itag or None if not found.
     get_lowest_resolution() → Optional[pytube.streams.Stream]
           Get lowest resolution stream that is a progressive mp4.
               Return type Stream or None
               Returns The Stream matching the given itag or None if not found.
     index (value | , start | , stop | | ) \rightarrow integer – return first index of value.
           Raises ValueError if the value is not present.
           Supporting start and stop arguments is optional, but recommended.
     last()
           Get the last Stream in the results.
               Return type Stream or None
               Returns Return the last result of this query or None if the result doesn't contain any streams.
     order by (attribute name: str) \rightarrow pytube.query.StreamQuery
           Apply a sort order. Filters out stream the do not have the attribute.
               Parameters attribute_name (str) – The name of the attribute to sort by.
     otf (is\_otf: bool = False) \rightarrow pytube.query.StreamQuery
           Filter stream by OTF, useful if some streams have 404 URLs
               Parameters is_otf (bool) – Set to False to retrieve only non-OTF streams
               Return type StreamQuery
               Returns A StreamQuery object with otf filtered streams
3.1.5 Caption Object
class pytube.Caption(caption_track: Dict[KT, VT])
     Container for caption tracks.
     download (title: str, srt: bool = True, output_path: Optional[str] = None, filename_prefix: Op-
                  tional[str] = None) \rightarrow str
           Write the media stream to disk.
```

Parameters itag(*int*) – YouTube format identifier code.

Parameters

- **title** (str) Output filename (stem only) for writing media file. If one is not specified, the default filename is used.
- **srt** Set to True to download srt, false to download xml. Defaults to True.

:type srt bool :param output_path:

(optional) Output path for writing media file. If one is not specified, defaults to the current working directory.

Parameters filename_prefix (str or None) – (optional) A string that will be prepended to the filename. For example a number in a playlist or the name of a series. If one is not specified, nothing will be prepended This is separate from filename so you can use the default filename but still add a prefix.

Return type str

```
static float\_to\_srt\_time\_format(d: float) \rightarrow str
```

Convert decimal durations into proper srt format.

Return type str

Returns SubRip Subtitle (str) formatted time duration.

float_to_srt_time_format(3.89) -> '00:00:03,890'

```
generate\_srt\_captions() \rightarrow str
```

Generate "SubRip Subtitle" captions.

Takes the xml captions from xml_captions () and recompiles them into the "SubRip Subtitle" format.

```
xml\_caption\_to\_srt(xml\_captions: str) \rightarrow str
```

Convert xml caption tracks to "SubRip Subtitle (srt)".

Parameters xml_captions (str) – XML formatted caption tracks.

xml_captions

Download the xml caption tracks.

3.1.6 CaptionQuery Object

```
class pytube.query.CaptionQuery (captions: List[pytube.captions.Caption])
```

Interface for querying the available captions.

all () \rightarrow List[pytube.captions.Caption]

Get all the results represented by this query as a list.

Return type list

get $(k[,d]) \rightarrow D[k]$ if k in D, else d. d defaults to None.

get_by_language_code ($lang_code: str$) \rightarrow Optional[pytube.captions.Caption] Get the Caption for a given lang_code.

Parameters lang_code (str) – The code that identifies the caption language.

Return type Caption or None

Returns The Caption matching the given lang_code or None if it does not exist.

items () \rightarrow a set-like object providing a view on D's items

```
keys () \rightarrow a set-like object providing a view on D's keys
values () \rightarrow an object providing a view on D's values
```

3.1.7 Extract

This module contains all non-cipher related data extraction logic.

```
pytube.extract.apply_descrambler(stream_data: Dict[KT, VT], key: str) \rightarrow None
     Apply various in-place transforms to YouTube's media stream data.
```

Creates a list of dictionaries by string splitting on commas, then taking each list item, parsing it as a query string, converting it to a dict and unquoting the value.

Parameters

- **stream_data** (dict) Dictionary containing query string encoded values.
- **key** (str) Name of the key in dictionary.

Example:

```
>>> d = {'foo': 'bar=1&var=test,em=5&t=url%20encoded'}
>>> apply_descrambler(d, 'foo')
>>> print(d)
{'foo': [{'bar': '1', 'var': 'test'}, {'em': '5', 't': 'url encoded'}]}
```

pytube.extract.apply_signature (config_args: Dict[KT, VT], fmt: str, js: str) \rightarrow None Apply the decrypted signature to the stream manifest.

Parameters

- config_args (dict) Details of the media streams available.
- fmt (str) Key in stream manifests (ytplayer_config) containing progressive download or adaptive streams (e.g.: url_encoded_fmt_stream_map or adaptive fmts).
- js(str) The contents of the base js asset file.

```
pytube.extract.get_ytcfg(html: str) \rightarrow str
      Get the entirety of the ytcfg object.
```

This is built over multiple pieces, so we have to find all matches and combine the dicts together.

Parameters html (str) – The html contents of the watch page.

Return type str

Returns Substring of the html containing the encoded manifest data.

```
pytube.extract.get_ytplayer_config(html: str) \rightarrow Any
```

Get the YouTube player configuration data from the watch html.

Extract the ytplayer_config, which is json data embedded within the watch html and serves as the primary source of obtaining the stream manifest data.

Parameters html (str) – The html contents of the watch page.

Return type str

Returns Substring of the html containing the encoded manifest data.

```
pytube.extract.get ytplayer js(html: str) \rightarrow Any
     Get the YouTube player base JavaScript path.
```

```
:param str html The html contents of the watch page.
```

```
Return type str
           Returns Path to YouTube's base.js file.
pytube.extract.initial data(watch html: str) \rightarrow str
     Extract the ytInitialData json from the watch_html page.
     This mostly contains metadata necessary for rendering the page on-load, such as video information, copyright
     notices, etc.
      @param watch_html: Html of the watch page @return:
pytube.extract.initial_player_response(watch_html: str) → str
     Extract the ytInitialPlayerResponse json from the watch_html page.
     This mostly contains metadata necessary for rendering the page on-load, such as video information, copyright
     notices, etc.
      @param watch html: Html of the watch page @return:
pytube.extract.is_age_restricted(watch_html: str) → bool
     Check if content is age restricted.
           Parameters watch_html (str) – The html contents of the watch page.
           Return type bool
           Returns Whether or not the content is age restricted.
pytube.extract.is_private(watch_html)
     Check if content is private.
           Parameters watch_html (str) – The html contents of the watch page.
           Return type bool
           Returns Whether or not the content is private.
pytube.extract.is_region_blocked(watch\_html: str) \rightarrow bool
     Determine if a video is not available in the user's region.
           Parameters watch_html (str) – The html contents of the watch page.
           Return type bool
           Returns True if the video is blocked in the users region. False if not, or if unknown.
pytube.extract.js url (html: str) \rightarrow str
     Get the base JavaScript url.
     Construct the base JavaScript url, which contains the decipher "transforms".
           Parameters html (str) – The html contents of the watch page.
```

 $pytube.extract.metadata(initial_data) \rightarrow Optional[pytube.metadata.YouTubeMetadata]$

{ 'Song': '(Gangnam Style)', 'Artist': 'PSY', 'Album': 'PSY SIX RULES Pt.1', 'Licensed to

Get the informational metadata for the video.

YouTube by': 'YG Entertainment Inc. [...]'

e.g.: [

1

Return type YouTubeMetadata

```
pytube.extract.mime_type_codec (mime_type_codec: str) \rightarrow Tuple[str, List[str]] Parse the type data.
```

Breaks up the data in the type key of the manifest, which contains the mime type and codecs serialized together, and splits them into separate elements.

Example:

```
mime_type_codec('audio/webm; codecs="opus"') -> ('audio/webm', ['opus'])
```

Parameters mime_type_codec (str) – String containing mime type and codecs.

Return type tuple

Returns The mime type and a list of codecs.

```
pytube.extract.playability_status (watch_html: str) -> (<class 'str'>, <class 'str'>)

Return the playability status and status explanation of a video.
```

For example, a video may have a status of LOGIN_REQUIRED, and an explanation of "This is a private video. Please sign in to verify that you may see it."

This explanation is what gets incorporated into the media player overlay.

Parameters watch_html (str) – The html contents of the watch page.

Return type bool

Returns Playability status and reason of the video.

```
pytube.extract.playlist_id(url: str) \rightarrow str Extract the playlist_id from a YouTube url.
```

This function supports the following patterns:

- https://youtube.com/playlist?list=playlist_id
- https://youtube.com/watch?v=video_id&list=playlist_id

Parameters url (str) – A YouTube url containing a playlist id.

Return type str

Returns YouTube playlist id.

```
pytube.extract.publish_date(watch_html: str)
```

Extract publish date :param str watch_html:

The html contents of the watch page.

Return type str

Returns Publish date of the video.

```
pytube.extract.recording_available(watch_html)
```

Check if live stream recording is available.

Parameters watch_html (str) – The html contents of the watch page.

Return type bool

Returns Whether or not the content is private.

```
pytube.extract.video_id (url: str) \rightarrow str
Extract the video id from a YouTube url.
```

This function supports the following patterns:

- https://youtube.com/watch?v=video_id
- https://youtube.com/embed/video_id
- https://youtu.be/video_id

Parameters url (str) – A YouTube url containing a video id.

Return type str

Returns YouTube video id.

pytube.extract.video_info_url ($video_id: str, watch_url: str$) \rightarrow str Construct the video_info url.

Parameters

- **video_id** (str) A YouTube video identifier.
- watch_url (str) A YouTube watch url.

Return type str

Returns https://youtube.com/get_video_info with necessary GET parameters.

pytube.extract.video_info_url_age_restricted($video_id: str, embed_html: str$) \rightarrow str Construct the video info url.

Parameters

- video_id (str) A YouTube video identifier.
- **embed_html** (str) The html contents of the embed page (for age restricted videos).

Return type str

Returns https://youtube.com/get_video_info with necessary GET parameters.

3.1.8 Cipher

This module contains all logic necessary to decipher the signature.

YouTube's strategy to restrict downloading videos is to send a ciphered version of the signature to the client, along with the decryption algorithm obfuscated in JavaScript. For the clients to play the videos, JavaScript must take the ciphered version, cycle it through a series of "transform functions," and then signs the media URL with the output.

This module is responsible for (1) finding and extracting those "transform functions" (2) maps them to Python equivalents and (3) taking the ciphered signature and decoding it.

```
\texttt{pytube.cipher.get\_initial\_function\_name}\ (\textit{js: str})\ \to \texttt{str}
```

Extract the name of the function responsible for computing the signature. :param str js:

The contents of the base.js asset file.

Return type str

Returns Function name from regex match

```
\texttt{pytube.cipher.get\_transform\_map} (\textit{js: str, var: str}) \rightarrow \texttt{Dict[KT, VT]}
```

Build a transform function lookup.

Build a lookup table of obfuscated JavaScript function names to the Python equivalents.

Parameters

- js(str) The contents of the base.js asset file.
- var (str) The obfuscated variable name that stores an object with all functions that descrambles the signature.

```
pytube.cipher.get_transform_object (js: str, var: str) \rightarrow List[str] Extract the "transform object".
```

The "transform object" contains the function definitions referenced in the "transform plan". The var argument is the obfuscated variable name which contains these functions, for example, given the function call DE.AJ (a, 15) returned by the transform plan, "DE" would be the var.

Parameters

- js(str) The contents of the base.js asset file.
- **var** (str) The obfuscated variable name that stores an object with all functions that descrambles the signature.

Example:

```
>>> get_transform_object(js, 'DE')
['AJ:function(a) {a.reverse()}',
'VR:function(a,b) {a.splice(0,b)}',
'kT:function(a,b) {var c=a[0];a[0]=a[b%a.length];a[b]=c}']
```

```
\texttt{pytube.cipher.get\_transform\_plan} (\textit{js: str}) \rightarrow List[str]
```

Extract the "transform plan".

The "transform plan" is the functions that the ciphered signature is cycled through to obtain the actual signature.

Parameters js(str) – The contents of the base js asset file.

Example:

```
['DE.AJ(a,15)', 'DE.VR(a,3)', 'DE.VR(a,51)', 'DE.KT(a,51)', 'DE.KT(a,51)', 'DE.KT(a,8)', 'DE.KT(a,21)']
```

```
pytube.cipher.map_functions(js\_func: str) \rightarrow Callable
```

For a given JavaScript transform function, return the Python equivalent.

Parameters $js_func(str)$ – The JavaScript version of the transform function.

```
pytube.cipher.reverse(arr: List[T], _: Optional[Any])
```

Reverse elements in a list.

This function is equivalent to:

```
function(a, b) { a.reverse() }
```

This method takes an unused b variable as their transform functions universally sent two arguments.

Example:

```
>>> reverse([1, 2, 3, 4])
[4, 3, 2, 1]
```

```
pytube.cipher.splice(arr: List[T], b: int)
```

Add/remove items to/from a list.

This function is equivalent to:

```
function(a, b) { a.splice(0, b) }
```

Example:

```
>>> splice([1, 2, 3, 4], 2)
[1, 2]
```

```
pytube.cipher.swap(arr: List[T], b: int)
```

Swap positions at b modulus the list length.

This function is equivalent to:

```
function(a, b) { var c=a[0];a[0]=a[b%a.length];a[b]=c }
```

Example:

```
>>> swap([1, 2, 3, 4], 2)
[3, 2, 1, 4]
```

3.1.9 Exceptions

Library specific exception definitions.

```
exception pytube.exceptions.ExtractError
```

Data extraction based exception.

```
exception pytube.exceptions.HTMLParseError
```

HTML could not be parsed

```
exception pytube.exceptions.LiveStreamError(video_id: str)
```

Video is a live stream.

```
exception pytube.exceptions.MaxRetriesExceeded
```

Maximum number of retries exceeded.

```
exception pytube.exceptions.MembersOnly(video_id: str)
```

Video is members-only.

YouTube has special videos that are only viewable to users who have subscribed to a content creator. ref: https://support.google.com/youtube/answer/7544492?hl=en

```
exception pytube.exceptions.PytubeError
```

Base pytube exception that all others inherit.

This is done to not pollute the built-in exceptions, which *could* result in unintended errors being unexpectedly and incorrectly handled within implementers code.

```
exception pytube.exceptions.RecordingUnavailable(video_id: str)
```

```
exception pytube.exceptions.RegexMatchError(caller: str, pattern: Union[str, Pattern[AnyStr]])
```

Regex pattern did not return any matches.

```
exception pytube.exceptions.VideoPrivate(video_id: str)
```

```
 \textbf{exception} \  \, \texttt{pytube.exceptions.VideoRegionBlocked} \, (\textit{video\_id: str}) \\
```

exception pytube.exceptions.**VideoUnavailable**(*video_id: str*)

Base video unavailable error.

3.1.10 Helpers

Various helper functions implemented by pytube.

```
\label{eq:cache}  \mbox{pytube.helpers.cache} \ (\mbox{\it func: Callable[[...], GenericType]}) \rightarrow \mbox{GenericType} \\  \mbox{mypy compatible annotation wrapper for lru_cache}
```

```
pytube.helpers.create_mock_html_json(vid\_id) \rightarrow Dict[str, Any]
```

Generate a json.gz file with sample html responses.

:param str vid_id YouTube video id

:return dict data Dict used to generate the json.gz file

```
pytube.helpers.deprecated(reason: str) \rightarrow Callable
```

This is a decorator which can be used to mark functions as deprecated. It will result in a warning being emitted when the function is used.

pytube.helpers.regex_search (pattern: str, string: str, group: int) \rightarrow str Shortcut method to search a string for a given pattern.

Parameters

- pattern (str) A regular expression pattern.
- **string** (*str*) A target string to search.
- **group** (*int*) Index of group to return.

Return type str or tuple

Returns Substring pattern matches.

```
pytube.helpers.safe_filename (s: str, max_length: int = 255) \rightarrow str Sanitize a string making it safe to use as a filename.
```

This function was based off the limitations outlined here: https://en.wikipedia.org/wiki/Filename.

Parameters

- **s** (str) A string to make safe for use as a file name.
- max_length (int) The maximum filename character length.

Return type str

Returns A sanitized string.

pytube.helpers.setup_logger(level: int = 40, log_filename: Optional[str] = None) \rightarrow None Create a configured instance of logger.

Parameters level (int) – Describe the severity level of the logs to handle.

```
pytube.helpers.target_directory(output_path: Optional[str] = None) \rightarrow str
```

Function for determining target directory of a download. Returns an absolute path (if relative one given) or the current path (if none given). Makes directory if it does not exist.

Returns An absolute directory path as a string.

```
pytube.helpers.uniqueify (duped\_list: List[T]) \rightarrow List[T]
```

Remove duplicate items from a list, while maintaining list order.

:param List duped_list List to remove duplicates from

:return List result De-duplicated list

3.1.11 Request

```
Implements a simple wrapper around urlopen.
```

```
pytube.request.filesize
```

Fetch size in bytes of file at given URL

Parameters url (str) – The URL to get the size of

Returns int: size in bytes of remote file

pytube.request.get (url, extra_headers=None, timeout=<object object>)
 Send an http GET request.

Parameters

- **url** (*str*) The URL to perform the GET request for.
- extra_headers (dict) Extra headers to add to the request

Return type str

Returns UTF-8 encoded string of response

```
pytube.request.head(url)
```

Fetch headers returned http GET request.

Parameters url (str) – The URL to perform the GET request for.

Return type dict

Returns dictionary of lowercase headers

Parameters

- **url** (*str*) The URL to perform the POST request for.
- extra_headers (dict) Extra headers to add to the request
- data (dict) The data to send on the POST request

Return type str

Returns UTF-8 encoded string of response

```
pytube.request.seq_filesize
```

Fetch size in bytes of file at given URL from sequential requests

Parameters url (str) – The URL to get the size of

Returns int: size in bytes of remote file

```
pytube.request.seq_stream(url, timeout=<object object>, max_retries=0)
```

Read the response in sequence. :param str url: The URL to perform the GET request for. :rtype: Iterable[bytes]

```
pytube.request.stream(url, timeout=<object object>, max_retries=0)
```

Read the response in chunks. :param str url: The URL to perform the GET request for. :rtype: Iterable[bytes]

$\mathsf{CHAPTER}\, 4$

Indices and tables

- genindex
- modindex
- search

Python Module Index

р

```
pytube, 13
pytube.cipher, 24
pytube.exceptions, 26
pytube.extract, 21
pytube.helpers, 27
pytube.request, 28
```

32 Python Module Index

Index

A all() (pytube.query.CaptionQuery method), 20 all() (pytube.query.StreamQuery method), 17	<pre>float_to_srt_time_format() (pytube.Caption</pre>
apply_descrambler() (in module pytube.extract), 21	G
apply_signature() (in module pytube.extract), 21 asc() (pytube.query.StreamQuery method), 17 author (pytube.YouTube attribute), 13 C cache() (in module pytube.helpers), 27 Caption (class in pytube), 19 caption_tracks (pytube.YouTube attribute), 13 CaptionQuery (class in pytube.query), 20 captions (pytube.YouTube attribute), 13 check_availability() (pytube.YouTube method),	<pre>generate_srt_captions()</pre>
<pre>create_mock_html_json() (in module py- tube.helpers), 27</pre>	tube.cipher), 24 get_lowest_resolution() (py-
default_filename (pytube.Stream attribute), 15 deprecated() (in module pytube.helpers), 27 desc() (pytube.query.StreamQuery method), 17 description (pytube.YouTube attribute), 13 download() (pytube.Caption method), 19 download() (pytube.Stream method), 15	<pre>tube.query.StreamQuery method), 19 get_transform_map() (in module pytube.cipher),</pre>
ExtractError, 26	<pre>tube.extract), 21 get_ytplayer_js() (in module pytube.extract), 21</pre>
F	H
filesize (in module pytube.request), 28 filesize (pytube.Stream attribute), 16 filesize_approx (pytube.Stream attribute), 16 filter() (pytube.query.StreamQuery method), 18 first() (pytube.query.StreamQuery method), 18	head() (in module pytube.request), 28 HTMLParseError, 26 I includes_audio_track (pytube.Stream attribute),

16 includes_video_track (pytube.Stream attribute), 16 index() (pytube.contrib.playlist.Playlist method), 15 index() (pytube.query.StreamQuery method), 19 initial_data() (in module pytube.extract), 22 initial_player_response() (in module py-	pytube (module), 13 pytube.cipher (module), 24 pytube.exceptions (module), 26 pytube.extract (module), 21 pytube.helpers (module), 27 pytube.request (module), 28 PytubeError, 26		
<pre>tube.extract), 22 is_adaptive (pytube.Stream attribute), 16</pre>	R		
is_age_restricted() (in module pytube.extract), 22 is_private() (in module pytube.extract), 22 is_progressive (pytube.Stream attribute), 16 is_region_blocked() (in module pytube.extract), 22	rating (pytube.YouTube attribute), 14 recording_available() (in module pytube.extract), 23 RecordingUnavailable, 26 regex_search() (in module pytube.helpers), 27 RegexMatchError, 26		
items() (pytube.query.CaptionQuery method), 20	register_on_complete_callback() (py-tube.YouTube method), 14		
J js_url() (in module pytube.extract), 22	register_on_progress_callback() (py- tube.YouTube method), 14 reverse() (in module pytube.cipher), 25		
K	S		
keys () (pytube.query.CaptionQuery method), 20 keywords (pytube.YouTube attribute), 14	<pre>safe_filename() (in module pytube.helpers), 27 seq_filesize (in module pytube.request), 28</pre>		
L last() (pytube.query.StreamQuery method), 19 length (pytube.YouTube attribute), 14 LiveStreamError, 26	<pre>seq_stream() (in module pytube.request), 28 setup_logger() (in module pytube.helpers), 27 splice() (in module pytube.cipher), 25 Stream(class in pytube), 15 stream() (in module pytube.request), 28</pre>		
M	stream_to_buffer() (pytube.Stream method), 17		
map_functions() (in module pytube.cipher), 25 MaxRetriesExceeded, 26 MembersOnly, 26	StreamQuery (class in pytube.query), 17 streams (pytube.YouTube attribute), 14 swap() (in module pytube.cipher), 26		
metadata (pytube. YouTube attribute), 14 metadata() (in module pytube.extract), 22 mime_type_codec() (in module pytube.extract), 23	T target_directory() (in module pytube.helpers), 27		
0	thumbnail_url (pytube.YouTube attribute), 14		
on_complete() (pytube.Stream method), 16 on_progress() (pytube.Stream method), 17 order_by() (pytube.query.StreamQuery method), 19 otf() (pytube.query.StreamQuery method), 19	title (pytube.contrib.playlist.Playlist attribute), 15 title (pytube.Stream attribute), 17 title (pytube.YouTube attribute), 14 trimmed() (pytube.contrib.playlist.Playlist method), 15		
P	U		
<pre>parse_codecs() (pytube.Stream method), 17 playability_status() (in module pytube.extract),</pre>	uniqueify() (in module pytube.helpers), 27		
player_response (pytube.YouTube attribute), 14 Playlist (class in pytube.contrib.playlist), 15 playlist_id() (in module pytube.extract), 23 post() (in module pytube.request), 28 publish_date (pytube.YouTube attribute), 14 publish_date() (in module pytube.extract), 23	<pre>values() (pytube.query.CaptionQuery method), 21 vid_info (pytube.YouTube attribute), 14 video_id() (in module pytube.extract), 23 video_info_url() (in module pytube.extract), 24 video_info_url_age_restricted() (in module pytube.extract), 24</pre>		

34 Index

```
video_urls (pytube.contrib.playlist.Playlist attribute),
15
VideoPrivate, 26
VideoRegionBlocked, 26
videos (pytube.contrib.playlist.Playlist attribute), 15
VideoUnavailable, 26
views (pytube.YouTube attribute), 14

X
xml_caption_to_srt() (pytube.Caption method),
20
xml_captions (pytube.Caption attribute), 20

Y
YouTube (class in pytube), 13
```

Index 35