

# Package ‘youtubecaption’

October 14, 2022

**Title** Downloading YouTube Subtitle Transcription in a Tidy Tibble Data Frame

**Version** 1.0.0

**Description** Although there exist some R packages tailored for YouTube API (e.g., 'tuber'), downloading YouTube video subtitle (i.e., caption) in a tidy form has never been a low-hanging fruit. Using 'youtube-transcript-api Python package' under the hood, this R package provides users with a convenient way of parsing and converting a desired YouTube caption into a handy 'tibble' data\_frame object. Furthermore, users can easily save a desired YouTube caption data as a tidy Excel file without advanced programming background knowledge.

**SystemRequirements** Anaconda (<https://www.anaconda.com/download/>)

**License** GPL-3

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**URL** <https://github.com/jooyoungseo/youtubecaption>

**BugReports** <https://github.com/jooyoungseo/youtubecaption/issues>

**Imports** stringr, writexl, reticulate, purrr, magrittr, tibble, dplyr

**RoxygenNote** 7.1.0

**Suggests** knitr, rmarkdown, covr

**NeedsCompilation** no

**Author** JooYoung Seo [aut, cre] (<<https://orcid.org/0000-0002-4064-6012>>),  
Soyoung Choi [aut] (<<https://orcid.org/0000-0002-0998-3352>>)

**Maintainer** JooYoung Seo <[jooyoung@psu.edu](mailto:jooyoung@psu.edu)>

**Repository** CRAN

**Date/Publication** 2020-05-15 05:50:02 UTC

## R topics documented:

get\_caption . . . . . 2

---

`get_caption`*Obtaining YouTube video caption/subtitle in a tidy tibble form.*

---

### Description

Use this function for downloading a desired YouTube video caption in a tidy tibble data\_frame form and save it as an Excel file in your current working directory.

### Usage

```
get_caption(  
  url = NULL,  
  language = "en",  
  savexl = FALSE,  
  openxl = FALSE,  
  path = getwd()  
)
```

### Arguments

<code>url</code>	A string value for a single YouTube video link URL. A typical form should start with "https://www.youtube.com/watch?v=" followed by a unique video ID.
<code>language</code>	two-character language code for the video URL. Set to "en" (English) by default. You can change this to fit with your needs (e.g., "ko" for Korean, "de" for German, etc.).
<code>savexl</code>	A logical value for determining whether or not to save the obtained tidy YouTube caption data as an Excel file. The default is FALSE which does not save it as a file. If set to TRUE, a file named "YouTube_caption_videoID.xlsx" is saved in your specified directory (the default is your current working directory).
<code>openxl</code>	A logical value for determining whether or not to open, if any, the saved YouTube_caption Excel file in your working directory. The default is FALSE. TRUE works only when the preceding argument (i.e., savexl) is set to TRUE.
<code>path</code>	A character vector of full path names; the default corresponds to the working directory, <code>getwd</code> . Tilde expansion (see <a href="#">path.expand</a> ) is performed. Missing values will be ignored.

### Details

`get_caption`

See example below.

### Value

tibble (advanced data.frame) object for a YouTube video caption will be returned.

**Author(s)**

JooYoung Seo, <jooyoung@psu.edu>

Soyoung Choi, <sxc940@psu.edu>

**References**

<https://pypi.org/project/youtube-transcript-api/>

**Examples**

```
library(youtubecaption)
# Let's get the video caption out of Hadley Wickham's "You can't do data science in a GUI":
url <- "https://www.youtube.com/watch?v=cpbtcsGE00A"
caption <- get_caption(url)
caption

# Save the caption as an Excel file and open it right it away
## Changing path to temp for the demonstration purpose only:
get_caption(url = url, savexl = TRUE, openxl = TRUE, path = tempdir())
```

# Index

\* **get\_caption**  
    get\_caption, 2

get\_caption, 2  
getwd, 2

path.expand, 2