

Youtube_Scrapper

January 28, 2025

```
[1]: # Importing necessary libraries
from bs4 import BeautifulSoup # For parsing HTML content
from selenium import webdriver # For automating web browser interaction

# Setting up the web driver for Chrome browser
browser=webdriver.Chrome()

[3]: # Open the target Youtube Channels videos page
browser.get("https://www.youtube.com/@GeeksforGeeksVideos/videos")

[5]: # Parse the page source with BeautifulSoup
soup=BeautifulSoup(browser.page_source,'html.parser')

[74]: # Find the main container for videos
sp=soup.find('ytd-rich-grid-renderer')

[82]: # Extract individual video elements
xp=sp.find_all('ytd-rich-item-renderer')

[104]: # Initialize a list to store video data
data=[]

# Loop through each video element to extract relevant details
for xp in sp.find_all('ytd-rich-item-renderer'):

    # Extracting video title
    title=xp.find('a',class_='yt-simple-endpoint focus-on-expand style-scope_
↳ytd-rich-grid-media').text

    # Extracting video link
    video_link="https://www.youtube.com/"+xp.
↳find('a',class_='yt-simple-endpoint focus-on-expand style-scope_
↳ytd-rich-grid-media').get('href')

    # Extracting video duration
    duration=xp.find_all('span')[0].text.replace("\n","").strip()

    # Extracting video views
```

```

views=xp.find_all('span')[5].text

# Extracting video upload_time
upload_time=xp.find_all('span')[6].text

# Extracting video thumbnail link
thumbnail_link=sp.find('img').get('src').split('?')[0]

# Appending Extracted Data to a List
data.append([title,video_link,duration,views,upload_time,thumbnail_link])

```

```

[106]: # Verifying extracted data as a list
len(data)

```

[106]: 1981

```

[108]: # Storing Extracted data as a list
import pandas as pd
df=pd.
↳DataFrame(data,columns=['title','video_link','duration','views','upload_time','thumbnail_link'])

```

```

[110]: df

```

```

[110]:

```

	title \	video_link	duration	views \
0	What is Machine Learning? Lecture 1 Free V...			
1	A Day in the Life at IIT: Struggles & Successe...			
2	DSA was hard until i learned these 10 Patterns...			
3	You're missing out by not starting the Three 9...			
4	Which Programming Language should you learn in...			
...
1976	Length of shortest chain to reach a target wor...			
1977	Binary Search GeeksQuiz			
1978	Number of Triangles in an Undirected Graph G...			
1979	Write a program to print all permutations of a...			
1980	Longest Span with same Sum in two Binary array...			
0		https://www.youtube.com/watch?v=1PtHLVsG_8s	10:46	452 views
1		https://www.youtube.com/watch?v=Ioj_XA-_jIY	8:06	1.6K views
2		https://www.youtube.com/watch?v=aWjFPloGYmg	6:32	5.1K views
3		https://www.youtube.com/watch?v=TcmoOfSCAJQ	0:32	265K views
4		https://www.youtube.com/watch?v=qduULN56XGk	8:37	1.5K views
...
1976		https://www.youtube.com/watch?v=6pIC20wCm20	12:25	44K views
1977		https://www.youtube.com/watch?v=T2sFYY-fT5o	7:33	192K views
1978		https://www.youtube.com/watch?v=ChdNz1Ui1uc	8:44	Now playing

```
1979 https://www.youtube.com/watch?v=AfxHGnRtFac 11:52 491K views
1980 https://www.youtube.com/watch?v=xtfj4-r_Ahs 20:26 307K views
```

```

      upload_time      thumbnail_link
0      1 day ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
1      2 days ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
2      4 days ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
3      8 days ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
4      9 days ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
...      ...      ...
1976 8 years ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
1977 8 years ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
1978      https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
1979 9 years ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg
1980 9 years ago https://i.ytimg.com/vi/1PtHLVsG_8s/hqdefault.jpg

```

[1981 rows x 6 columns]

```
[112]: # Check for null values in a DataFrame
df.isnull().sum()
```

```
[112]: title      0
video_link      0
duration        0
views           0
upload_time     0
thumbnail_link  0
dtype: int64
```

```
[114]: # Exporting data to a CSV File
df.to_csv('Youtube_Scrapper.csv')
```

```
[547]: # Extracting detailed data for each video by visiting individual links
import time
import numpy as np
```

```
[549]: # Reinitializing Web Driver
browser=webdriver.Chrome()

browser.get('https://www.youtube.com/@GeeksforGeeksVideos/videos') # Opening
↳ Youtube Page of GFG
time.sleep(2) # Wait for the page to load
data=[]

# Looping through each video link to extract additional details
for link in df['video_link']:
```

```

browser.get(link) # Open the video link

time.sleep(5) # Wait for the page to load

# Parse the video page
soup=BeautifulSoup(browser.page_source, 'html.parser')

# Extracting additional details with error handling
try:
    title=soup.find('yt-formatted-string',class_='style-scope_
↪yt-d-watch-metadata').text
except:
    title=np.nan
try:
    view=soup.find('span',class_='bold style-scope yt-formatted-string').text
except:
    view=np.nan
try:
    upload_date=(soup.find_all('span',class_='bold style-scope_
↪yt-formatted-string'))[2].text
except:
    upload_date=np.nan
try:
    like=soup.
↪find_all(class_='yt-spec-button-shape-next__button-text-content')[10].text
except:
    like=np.nan
try:
    description=soup.
↪find(class_='yt-core-attributed-string--link-inherit-color').text
except:
    description=np.nan

# Appending extracted details
data.append([title,view,upload_date,like,description])

```

```

[551]: # Create a DataFrame with the detailed data
df=pd.
↪DataFrame(data,columns=['title','view','upload_date','like','description'])

```

```

[553]: # Display the detailed DataFrame
df

```

```
[553]:
```

		title	view \
0		What is Machine Learning? Lecture 1 Free V...	453 views
1		A Day in the Life at IIT: Struggles & Successe...	1.6K views
2		DSA was hard until i learned these 10 Patterns...	5.3K views
3		You're missing out by not starting the Three 9...	265K views
4		Which Programming Language should you learn in...	1.5K views
...	
1976		Length of shortest chain to reach a target wor...	44K views
1977		Binary Search GeeksQuiz	192K views
1978		Number of Triangles in an Undirected Graph G...	19K views
1979		Write a program to print all permutations of a...	491K views
1980		Longest Span with same Sum in two Binary array...	307K views

	upload_date	like	description
0	1 day ago	About	Welcome to the first video of GeeksforGeeks' c...
1	2 days ago	About	Watch Full Video: -
2	4 days ago	Videos	These 10 Patterns Are All You Need to Solve DS...
3	8 days ago	About	The best version of you is just 90 days away \...
4	9 days ago	Videos	Start the challenge now:
...
1976	8 years ago	About	Explanation for the article:
1977	8 years ago	About	Explanation for the article:
1978	8 years ago	Videos	Explanation for the article:
1979	9 years ago	About	Explanation for the Article:
1980	9 years ago	About	Here's you next clue - Our comprehensive guid...

[1981 rows x 5 columns]

```
[555]: # Check for null values in the detailed DataFrame
df.isnull().sum()
```

```
[555]: title          119
view            119
upload_date     119
like            119
description     118
dtype: int64
```

```
[557]: # Checking length of Data
len(data)
```

```
[557]: 1981
```

```
[559]: # Exporting the detailed data to a CSV file
df.to_csv("Video_Description.csv")
```

```
[ ]:
```