INTRODUCTION × BACKGROUND × APPLICATIONS × TOOLS × DEMONSTRATION ×

CMAA5200/EMIA6500F (Spring 2024) – Tutorial Session: Christie Pang

APIs-Based Research

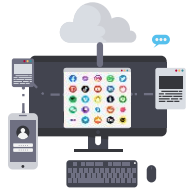
Using APIs for Data Collection on Social Media Research

APIs

Why do we analyze social media?

Pervasive Social Media

Enabling communication among users and collecting massive amounts of data for business and research



Social Media Research

Becomes a crucial field in research (HCI / computational social sciences / digital sociology)

Social Media Research → Qualitative Methods → Usage Patterns

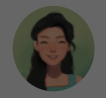
i.e., in-depth interview, survey, semi-structured interview …

Social Media Research → Systematic Approach to access usage patterns

A collection of detailed online information on large populations → analysis → How?

INTRODUCTION × BACKGROUND × APPLICATIONS × TOOLS × DEMONSTRATION ×

Using APIs for Data Collection on Social Media Research: Background and Overview



APIs

Why do we analyze social media?

Pervasive Social Media

Enabling communication among users and collecting massive amounts of data for business and research

WHAT CAN WE DO?

Social Media Research

Becomes a crucial field in research (HCI / computational social sciences / digital sociology)

Social Media Research → Qualitative Methods → Usage Patterns

i.e., in-depth interview, survey, semi-structured interview …

Social Media Research → Systematic Approach to access usage patterns

A collection of detailed online information on large populations → analysis → How?

APIs

Definition:

A back-end interface through which third-party developers may connect new add-ons to an existing service

In the Past…

British computer scientists Maurice Wilkes and David Wheeler worked on the modular software library in 1940s for EDSAC (early computer); Wikes and Wheeler (1951) specified API in *The Preparation of Programs for an Electronic Digital Computer*

Now…

Many social media companies make their data banks on users and usage patterns available through their APIs → USEFUL TOOL

APIs

Pros…

- Automated method → efficient
- Internet research → allows us to gather individual's digital footprints → access to usage patterns *(i.e., communication and connectivity…)* in a systematic and detailed way
- **Web crawling** (traverse only the public web) VS **APIs** (provide access to on-public internet environment)
- Use of **Raw Log Files** (may need some official collaboration) VS **APIs** (do not require formalized collaboration or affiliation)

Cons…

- We don't know to what extent the social media APIs are "open" to us
- A lack of transparency in data output and quality
- Economical consideration…
- Sudden (more and more) restriction…

APIs

How can we use APIs as a Research Tool?

User Behavior Analysis & Interaction Patterns

Quantitative Research

- Harvesting and analyzing so-called “big data”
- In terms of pattern recognition and detection of deviant patterns that need further attention

Qualitative Research

- Harvesting textual archives of communication patterns on social media for closeup analysis

1 Content Analysis: Sentiment, Behaviour, Emotions

2 Relational Data and Network Analysis

3 Systems based on Social Media Data

4 Sampling Tool for Survey-based Studies

APIs

How can we use APIs as a Research Tool?

Case Studies and Related Works

1 Content Analysis: Sentiment, Behaviour, Emotions

Close-up and Whispering: An Understanding of Multimodal and Parasocial Interactions in YouTube ASMR videos
(Niu, Manon, Bartolome, Ha, Veazey, 2022)

2 Relational Data and Network Analysis

Social Networks that Matter: Twitter Under the Microscope
(Huberman, Romero, Wu, 2009)

Who Says What to Whom on Twitter
(Wu, Hofman, Mason, Watts, 2011)

3 Systems based on Social Media Data

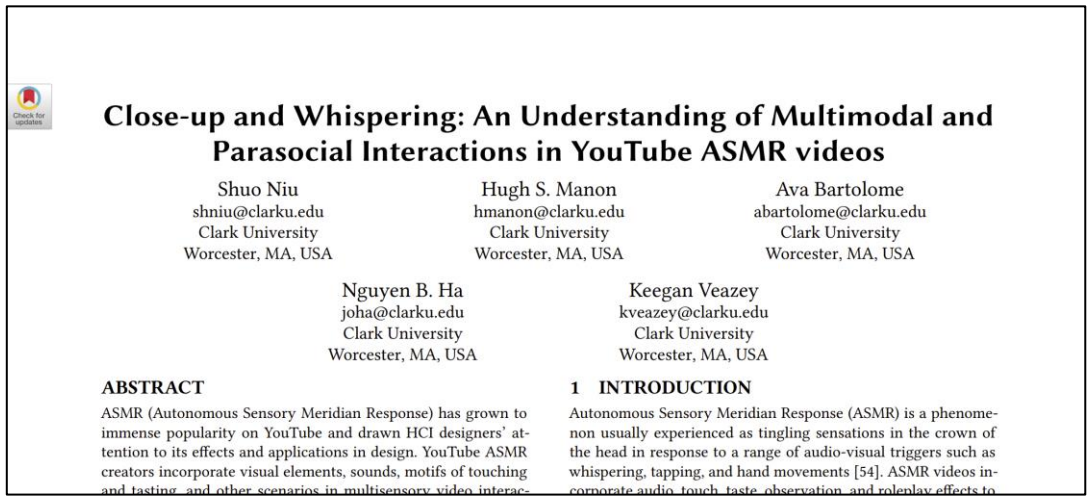
Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach
(Kowald, Pujari, Lex, 2017)

4 Sampling Tool for Survey-based Studies

Teenage uploaders on YouTube: Networked public expectancies, online feedback preference, and received on-platform feedback
(Courtois, Mechant, Marez, 2011)

APIs

How can we use APIs as a Research Tool?



Close-up and Whispering: An Understanding of Multimodal and Parasocial Interactions in YouTube ASMR videos

Shuo Niu
shniu@clarku.edu
Clark University
Worcester, MA, USA

Hugh S. Manon
hmanon@clarku.edu
Clark University
Worcester, MA, USA

Ava Bartolome
abartolome@clarku.edu
Clark University
Worcester, MA, USA

Nguyen B. Ha
joha@clarku.edu
Clark University
Worcester, MA, USA

Keegan Veazey
kveazey@clarku.edu
Clark University
Worcester, MA, USA

ABSTRACT

ASMR (Autonomous Sensory Meridian Response) has grown to immense popularity on YouTube and drawn HCI designers' attention to its effects and applications in design. YouTube ASMR creators incorporate visual elements, sounds, motifs of touching and tasting, and other scenarios in multisensory video interac-

1 INTRODUCTION

Autonomous Sensory Meridian Response (ASMR) is a phenomenon usually experienced as tingling sensations in the crown of the head in response to a range of audio-visual triggers such as whispering, tapping, and hand movements [54]. ASMR videos incorporate audio, touch, taste, observation, and roleplay effects to

Close-up and Whispering: An Understanding of Multimodal and Parasocial Interactions in YouTube ASMR videos – link (Niu, Manon, Bartolome, Ha, Veazey, 2022)

Overview

- **ASMR** (Autonomous Sensory Meridian Response)
- Immense popularity on YouTube and research has identified the mental wellbeing → understudied phenomenon in HCI community + constrain designers' ability to incorporate ASMR in video-based designs
- This work annotates how ASMRtists use different senses to deliver **attractiveness** → obtains the distribution of ASMR interaction modalities and parasocial attractions reveal patterns of ASMR experiences
- Multimodal social connection, relaxing physical intimacy, and sensory-rich activity observation
- A key finding is **the feeling-oriented words** from viewer comments

APIs

How can we use APIs as a Research Tool?

Method (Collecting ASMR Video Data)

227,133 videos

- YouTube Data API
- Search seed "ASMR"
- Jun 11 & Oct 20, 2020 + Posted in prior 3 months = active channel

300 Comments Per Video

- Return 300 top-level comments for each video
- Ensure similar amount of comments regardless popularity

Filtering

- Non-English video → difficulties in data tagging and categorization
- <5 minutes (N=9676)
- Time range: Jan 01 to Jun 01
- < 50 comments (31.39%)

85,734 Videos

- From 697 channels
- Randomly sample 200 for grounded theory analysis

APIs

What are some helpful tools?

Official APIs

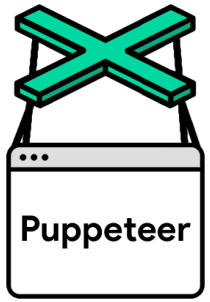
- Meta for Developers (link)
  - Instagram Graph API, YouTube Data API v3
- Twitter API (X Developer)
- TikTok API (link)
- Xiaohongshu API (link)
- Weibo API (link)
- And more…

Existing Dataset & Sources

- Harvard Dataverse (home)
- SNAP (link)
- SMILE (home)
- Catalog (home)
- Data for Good at Meta (link)
- TweetSets (home) & Tutorial
- Kaggle (home)
- CrowdTangle (link)
- SOMAR @ ICPSR (link)

Open-Source Libraries

- Puppeteer (Docs, Chinese version)
- Selenium (Docs)
- Playwright (Docs)



Others

- Social Media Research Toolkit
- YouTube Data Tools
- SAGE List of Social Media Tools
- TikTok API SDK (link, tutorial)

APIs

How can we scrap YouTube Videos according to keywords & pull comments from multiple videos?

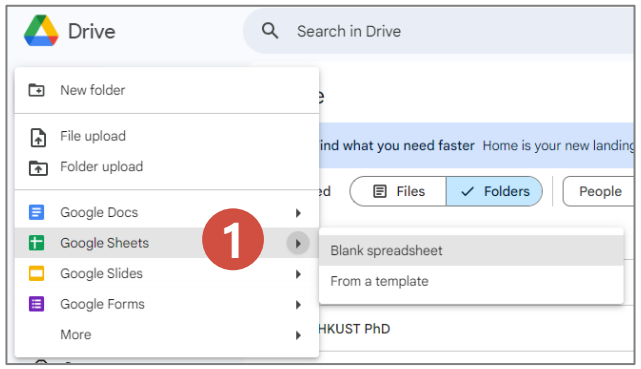
To cater to students from different disciplines and backgrounds, this demonstration does not require any specific downloads and can be used in different environment (if it fits your practice, and you are familiar with that)!

https://colab.research.google.com/drive/1DD8n2vZj279IMDEunIEV9L4ZTHIHS1vY?usp=sharing

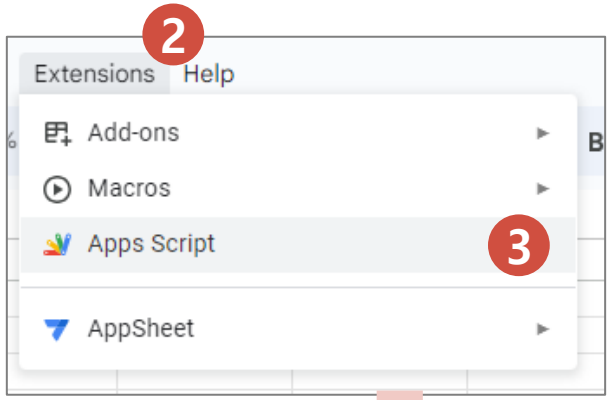
APIs

How can we scrap YouTube Videos according to keywords & pull comments from multiple videos?

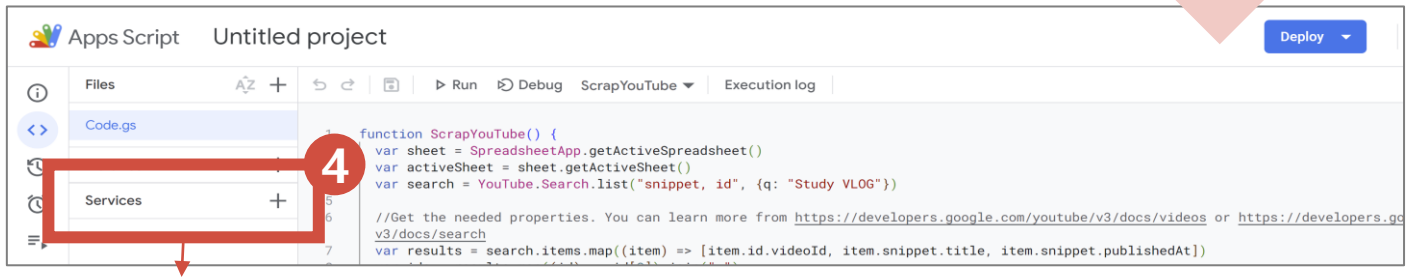
Scrap YT Video List



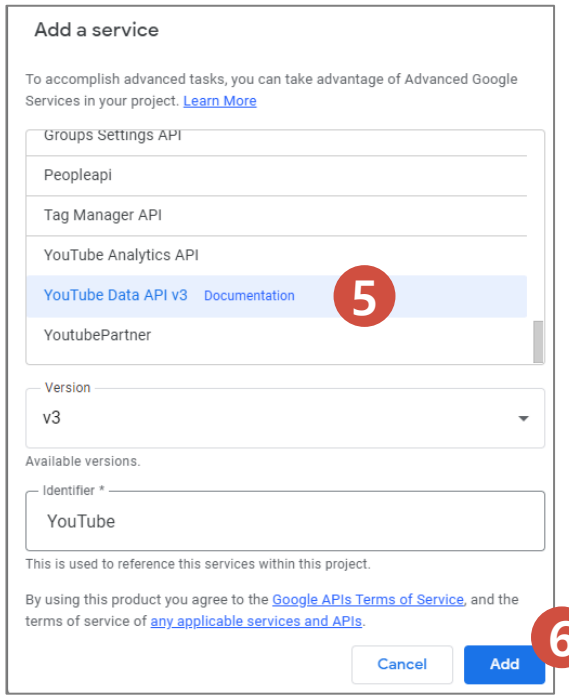
Open a Google Sheets in your Google Drive



Extensions → Apps Script



Add the Services in your project (API)



Add YouTube Data API and its identifier

APIs

How can we scrap YouTube Videos according to keywords & pull comments from multiple videos?

Scrap YT Video List

7

```
1  function ScrapYouTube() {
2    var sheet = SpreadsheetApp.getActiveSpreadsheet()
3    var activeSheet = sheet.getActiveSheet()
4    var search = YouTube.Search.list("snippet, id", {q: "YOUR KEYWORD HERE", maxResults: 50})
5
6    //Test whether the function works
7    Logger.log(search)
8  }
```

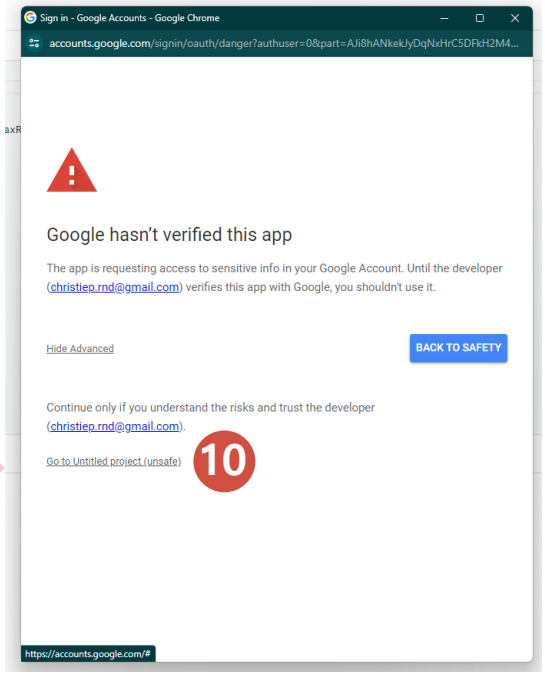
This function triggers the YouTube Data API (self-defined identifier. In this case, we use the default “YouTube”) and scrap 50 YouTube videos according to the inputted keyword. Click ▷ Run to execute.

8 9 ▷ Run

Execution log

4:38:41 PM Error Attempted to execute myFunction, but it was deleted.

You will receive the error in execution log for the first run. Let’s click ▷ Run again.



Go to the unsafe mode and authorize with your Google account.

APIs

How can we scrap YouTube Videos according to keywords & pull comments from multiple videos?

Scrap YT Video List

11

```
//Get the needed properties. You can learn more from https://developers.google.com/youtube/v3/docs/videos or https://developers.google.com/youtube/v3/docs/search
var results = search.items.map((item) => [item.id.videoId, item.snippet.title, item.snippet.publishedAt])
var ids = results.map((id) => id[0]).join(",")
var stats = YouTube.Videos.list("statistics", {id: ids})
var videoStats = stats.items.map((item) => [item.statistics.viewCount, item.statistics.likeCount, item.statistics.commentCount])
```

12

```
//Extract the data to Google Sheet
activeSheet.getRange(2, 1, results.length, results[0].length).setValues(results)
activeSheet.getRange(2, 4, videoStats.length, results[0].length).setValues(videoStats)
```

Get all properties that cater to your project. For details for what you can choose via YouTube Data VPI, please refer to https://developers.google.com/youtube/v3/docs/search or https://developers.google.com/youtube/v3/docs/videos . Then, it is time to extract the data to your current sheet.

13

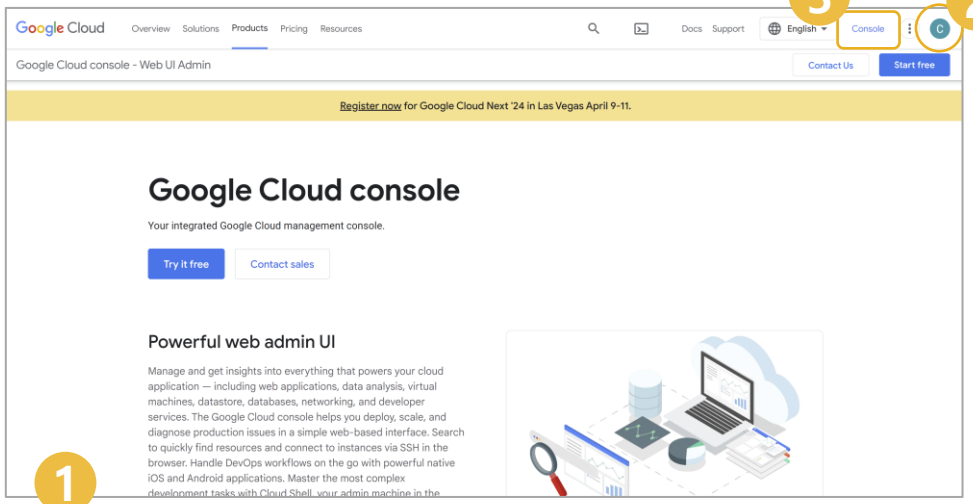
	A	B	C	D	E	F
1	ID	Title	Publish At (Time)	View	Like	Comment
2	iGVALUOKKbY	SOCIAL MEDIA - UCHE MONTANA OKUSAGA ADEOLUWA EBERE OKAFOR	2023-12-28T16:08:20Z	3070933	62423	7496
3	oDihnbzhRVc	Social Media Vs. Reality: Hotel Edition! #travel #shorts #socialmedia #reality	2023-12-21T15:15:03Z	10328832	141975	527
4	chMRLqlPs50	Worst Company Social Media Mistakes	2023-07-30T17:15:02Z	5963751	501789	4849
5	VDmeGQcpRLQ	WATCH LIVE: CEOs of Meta, TikTok, X and other social media companies testify in Senate hearing	2024-01-31T19:03:58Z	195737	2448	
6	5SwSb3k2LPM	ALLMO$T - Social Media (Damn) [Official Music Video]	2020-03-28T05:00:12Z	2559297	24706	1333

Now, your data is ready on the sheet! You can also add table headers and sorting functions.

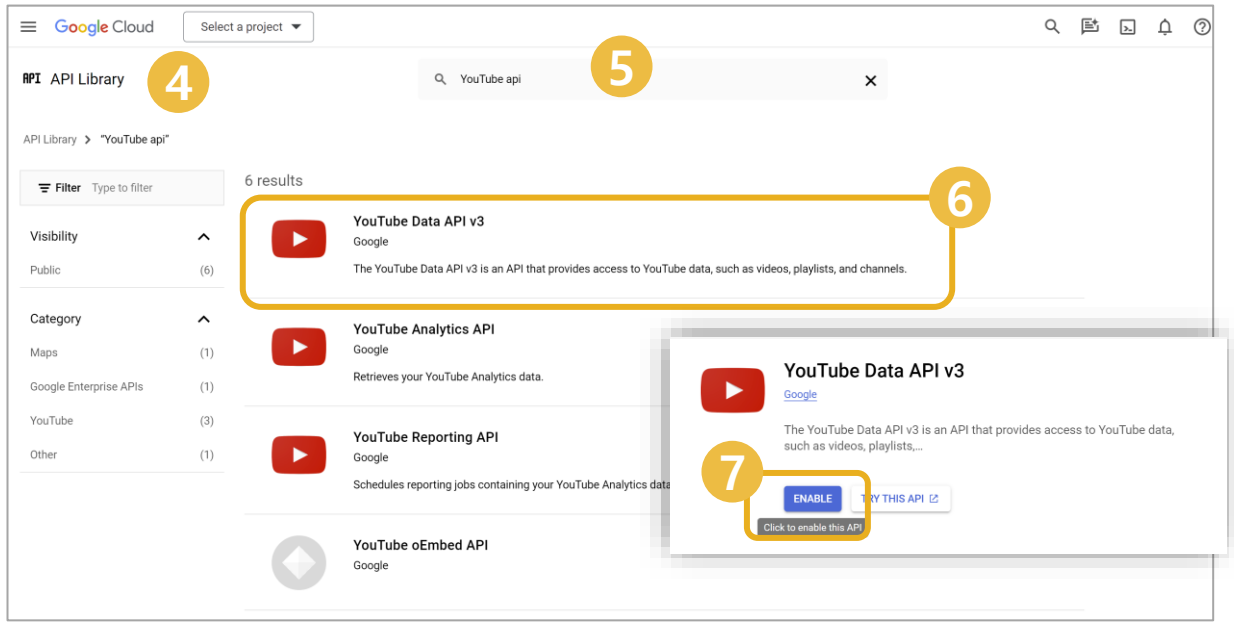
APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Get Your API



Go to https://cloud.google.com/?hl=en, sign in and direct to own Console.

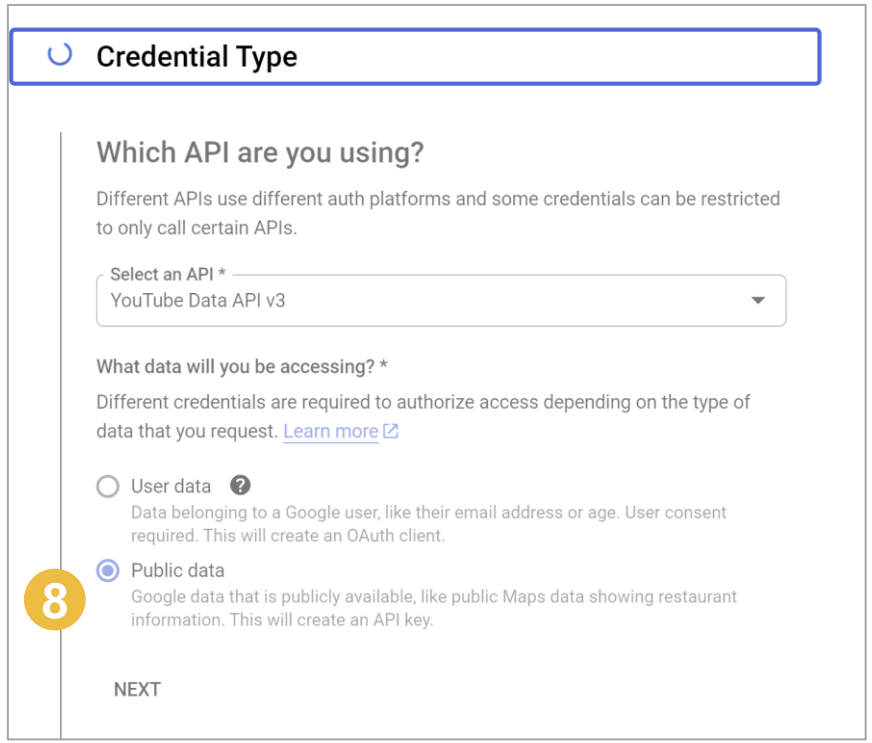


Go to your API Library, search and enable the YouTube Data API v3.

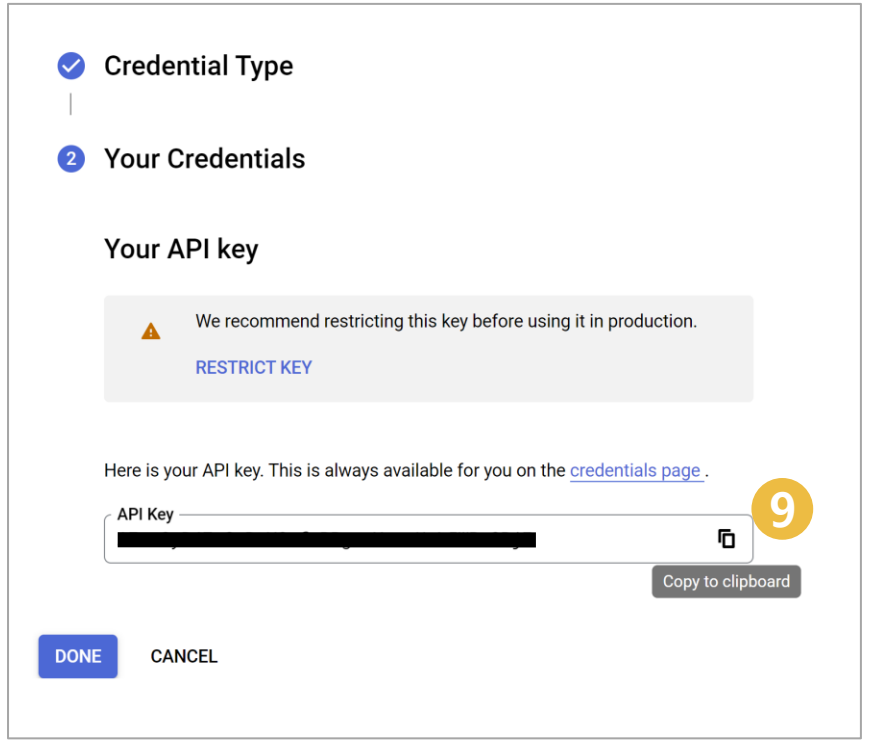
APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Get Your API



Enable your API key.



Remember to COPY your key. This serves as an ID for you! Please keep it hidden and anonymous.

APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Scrap Comments I

Use whatever platform / environment that you are used to. In this case, to make sure this tutorial is suitable for all students including novice, we will use Google colab and Python.

```
key = "HIDDEN - place your key here"

# Basic Information and Environment
import googleapiclient.discovery
import pandas as pd

api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY = key

youtube = googleapiclient.discovery.build(
    api_service_name, api_version, developerKey=DEVELOPER_KEY
)

# Set a function: Get Comments (video id)
def getcomments(video):
  request = youtube.commentThreads().list(
      part="snippet",
      videoId=video,
      maxResults=100
  )

comments = []
```

```
# Execute the request.
response = request.execute()

# Get the comments from the response. Look for the properties a
nd representation here on https://developers.google.com/youtube
/v3/docs/comments

# Note that not all comments can be extracted as it only pulls
the first level comments
for item in response['items']:
    comment = item['snippet']['topLevelComment']['snippet']
    public = item['snippet']['isPublic']
    comments.append([
        comment['authorDisplayName'],
        comment['publishedAt'],
        comment['likeCount'],
        comment['textOriginal'],
        comment['videoId'],
        public
    ])
```

APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Scrap Comments I

```
while (1 == 1):
    try:
     nextPageToken = response['nextPageToken']
    except KeyError:
     break
    nextPageToken = response['nextPageToken']
    # Create a new request object with the next page token.
    nextRequest = youtube.commentThreads().list(part="snippet", videoId=video, maxResults=100, pageToken=nextPageToken)
    # Execute the next request.
    response = nextRequest.execute()
    # Get the comments from the next response. Remember to change the properties if they were changed outside the while loop.
    for item in response['items']:
      comment = item['snippet']['topLevelComment']['snippet']
      public = item['snippet']['isPublic']
      comments.append([
          comment['authorDisplayName'],
          comment['publishedAt'],
          comment['likeCount'],
          comment['textOriginal'],
          comment['videoId'],
          public
      ])

  df2 = pd.DataFrame(comments, columns=['author', 'updated_at', 'like_count', 'text','video_id','public'])
  return df2
```

APIs

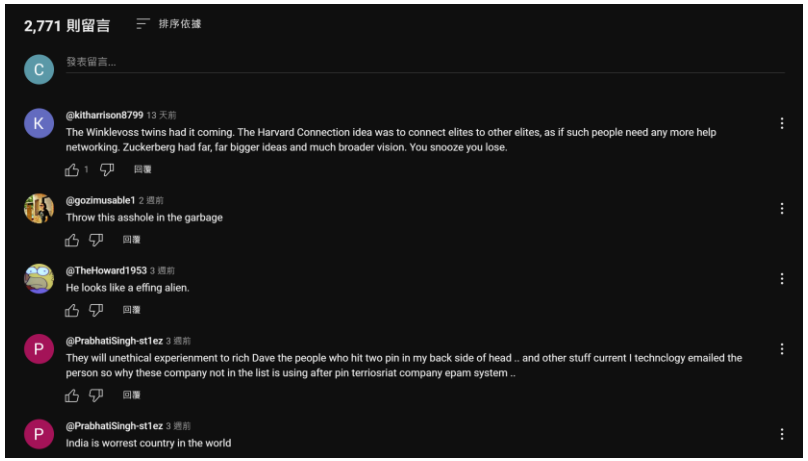
How to get our YouTube Data API / APIs via Google Cloud Platform?

Scrap Comments I

Use whatever platform / environment that you are used to. In this case, to make sure this tutorial is suitable for all students including novice, we will use Google colab and Python.

```
# Test the function with a single video
df = getcomments('VIDEO ID HERE')
df
```

Results:



	author	updated_at	like_count	text	video_id	public
0	@kitharrison8799	2024-01-28T08:21:50Z	1	The Winklevoss twins had it coming. The Harvar...	5WiDIhIkPoM	True
1	@gozimusable1	2024-01-28T05:01:26Z	0	Throw this asshole in the garbage	5WiDIhIkPoM	True
2	@TheHoward1953	2024-01-18T16:31:56Z	0	He looks like a effing alien.	5WiDIhIkPoM	True
3	@PrabhatiSingh-st1ez	2024-01-18T13:16:40Z	0	They will unethical experienment to rich Dave ...	5WiDIhIkPoM	True
4	@PrabhatiSingh-st1ez	2024-01-18T13:15:27Z	0	India is worrest country in the world	5WiDIhIkPoM	True
...	...	...	...	...	...	...
1863	@chrisc5197	2013-06-20T02:25:07Z	3	congress - attacks Mark Zuckerberg for invadin...	5WiDIhIkPoM	True
1864	@AntoniaVerastoninhaveras	2013-06-19T00:49:50Z	2	TAN SOLO TÚ	5WiDIhIkPoM	True
1865	@Zendket08	2013-06-08T02:39:33Z	0	es super guapo	5WiDIhIkPoM	True
1866	@Zendket08	2013-06-08T02:36:30Z	0	en este dia loqueo todos los chats	5WiDIhIkPoM	True
1867	@fragergamingdie4983	2013-06-06T11:03:29Z	0	but he created facebook by python,php,javascri...	5WiDIhIkPoM	True

1868 rows × 6 columns

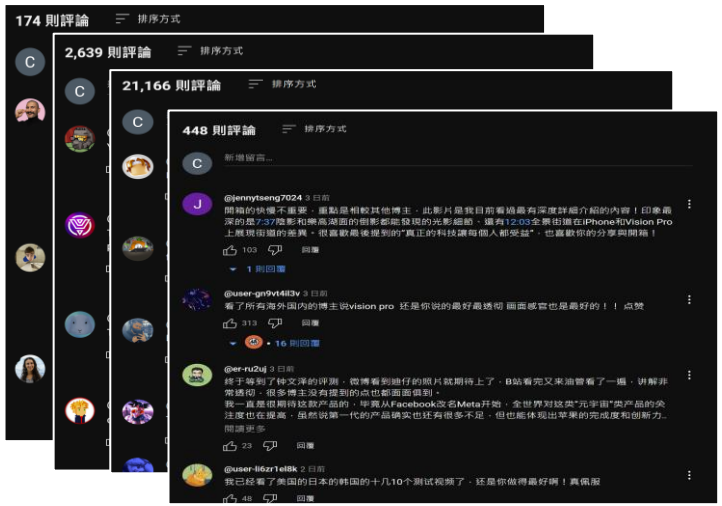
APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Scrap Comments II

```
# For Loop for Getting the Comments from a List of Video IDs
df = pd.DataFrame()
for i in [VIDEO ID TWO HERE', 'VIDEO ID THREE HERE','VIDEO ID FOUR HERE',AND SO ON']:
  df2 = getcomments(i)
  df = pd.concat([df, df2])
```

Results:



df

	author	updated_at	like_count	text	video_id	public
0	@Yorumcu666	2024-02-11T13:43:17Z	0	I don't really understand why it is a great th...	BV9Xy6L_rlM	True
1	@micacam2684	2024-02-11T13:27:50Z	0	Does it print?	BV9Xy6L_rlM	True
2	@JAYJAYfrdoe	2024-02-11T13:09:00Z	0	Yup the movies were right, the end is near	BV9Xy6L_rlM	True
3	@99mully99	2024-02-11T13:01:36Z	0	yep	BV9Xy6L_rlM	True
4	@namphanphunghoang1011	2024-02-11T12:44:52Z	0	They make us to be stick illusion more. One da...	BV9Xy6L_rlM	True
...	...	...	...	...	...	...
379	@anthonychen4052	2024-02-08T11:02:09Z	0	第二名	ya5vAbFhf6U	True
380	@yiantan	2024-02-08T11:01:44Z	1	終於啊	ya5vAbFhf6U	True
381	@user-rr3bh2su7d	2024-02-08T11:01:42Z	1	总算等到文泽的了	ya5vAbFhf6U	True
382	@edwardzhang6581	2024-02-08T11:01:33Z	3	快男，终于等到你	ya5vAbFhf6U	True
383	@ying-shanlin4190	2024-02-08T11:01:22Z	0	1	ya5vAbFhf6U	True

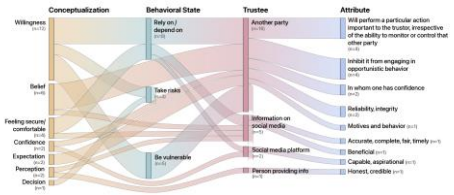
21071 rows × 6 columns

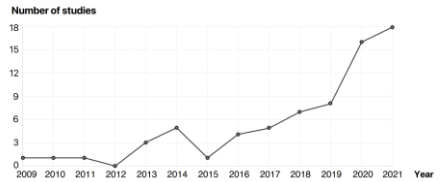
APIs

How to get our YouTube Data API / APIs via Google Cloud Platform?

Play More with It!

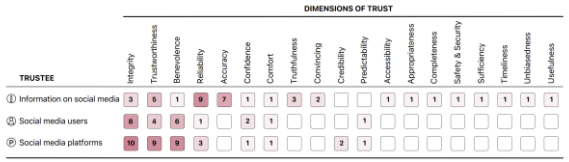
- After getting the list of comments, you can play with it on your environment to create different interesting and meaningful charts. Meanwhile, extracting it as JSON or CSV to further 'read' and 'design' your own visualization.
  - `df.to_csv(file.csv', index=False)`
  - Some popular visualization includes Sankey diagram, word cloud, time-series chart, traditional table formats, heat map, and more.





Platform	Number of studies	Percentage of studies*%
General	27	39%
Facebook	13	19%
Twitter	8	11%
WeChat	4	6%
Instagram	3	4%
Weibo	2	3%
YouTube	2	3%
Snapchat	2	3%
Others **	7	10%

*Total % is more than 100% as some papers used multiple methods.
** If a platform was only examined by one study, then this platform [illegible] "others".



*Reference: Yixuan Zhang, Joseph D Gaggiano, Nutchanon Yongsatianchot, Nurul M Suhaimi, Miso Kim, Yifan Sun, Jacqueline Griffin, and Andrea G Parker. 2023. What Do We Mean When We Talk about Trust in Social Media? A Systematic Review. In Proceedings of *the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 670, 1–22. https://doi-org.lib.ezproxy.hkust.edu.hk/10.1145/3544548.3581019

Kalmukov, Yordan. (2021). Using word clouds for fast identification of papers' subject domain and reviewers' competences.

APIs

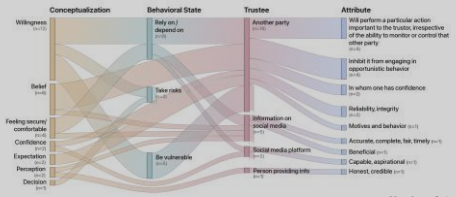
How to get our YouTube Data API / APIs via Google Cloud Platform?

Play More with It!

- After getting the list of c… t to create different interesting and meaning… to further 'read' and 'design' your own visualization.
  - `df.to_csv(file…`
  - Some popular visua… -series chart, traditional table formats, heat…

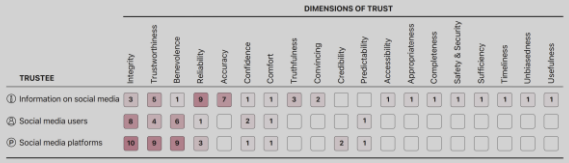
Q&A

Thank you



Platform	Number of studies	Percentage of studies*%
General	27	39%
Facebook	13	19%
Twitter	8	11%
WeChat	4	6%
Instagram	3	4%
Weibo	2	3%
YouTube	2	3%
Snapchat	2	3%
Others **	7	10%

*Total % is more than 100% as some papers used multiple methods.
** If a platform was only examined by one study, then this platform …"others".



*Reference: Yixuan Zhang, Joseph D Gaggiano, Nutchanon Yongsatianchot, Nurul M Suhaimi, Miso Kim, Yifan Sun, Jacqueline Griffin, and Andrea G Parker. 2023. What Do We Mean When We Talk about Trust in Social Media? A Systematic Review. In Proceedings of *the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 670, 1–22. https://doi-org.lib.ezproxy.hkust.edu.hk/10.1145/3544548.3581019
Kalmukov, Yordan. (2021). Using word clouds for fast identification of papers' subject domain and reviewers' competences.