

Codebook

Prepared by Sebastián Rojas Cabal

Last updated: November 15, 2018

SAMPLE OF WORK:

Please read this document and extract the relevant data for this video:

<https://www.youtube.com/watch?v=whGSgJuLMfQ>

Task description

We want to extract information from YouTube that will allow us to build a dataset connecting different right-wing political content producers and consumers in the US. The task consists of scraping the data from YouTube and processing it into a format we can use to construct the network database as well as analyze the content of their interactions.

Scraping data from YouTube

We ask you to familiarize yourself with how to scrape data from YouTube and how to the audio from the YouTube videos into text. For scraping we recommend using the [tuber package](#) for R. You can also explore ways of using the YouTube API. If you choose to use tuber and R, [this tutorial](#) is helpful in becoming familiar with the process of scraping data from YouTube.

Automated text transcription

YouTube provides time-stamped transcripts of its videos.

Variables

Below you will find a list of variables and their definitions, and some examples on how to extract and input data.

Table 1. List of variables

Variable	Description
video_id	The alphanumeric figures after the equality symbol (=) in the video's URL (see figure 1).
video_url	The complete video URL.
video_title	Title of the video posted, or under which the comments were made.
channel	Name of the channel on which the video was posted on.
series	Name of the series of which the video is a part of.
video_snippet	Text from the video description.
video	Takes the value of 1 if it is a video, 0 if otherwise.
comment	Takes the value of 1 if it is a comment, 0 if otherwise.

Variable	Description
reply	Takes the value of 1 if it is a reply, 0 if otherwise.
video_op	The name of the person/user posting the video.
comment_op	The name of the person/user posting the comment/for which the comment is a reply to. NA if the observation is not a comment or reply.
reply_op	The name of the person/user posting the reply. NA if the observation is not a reply.
date_posted	The date the video/comment/reply was posted. Format: YYYY_MM_DD_HH_MM_SS
content	A text transcript of the content of the video or the text of the comment/reply.
date_scrape	The date the video and related content was scraped. Format: YYYY_MM_DD_HH_MM_SS

Notes on the variables:

- We will provide you with a list of channel, series and influencer names. You will use this information to assign values to the ‘channel’, ‘series’, and ‘video_op’ variables.
- For dates of posting and scraping, use US Eastern Standard Time as the time zone when you extract the data. If you cannot adjust this setting, then specify the time zone you used.
 - If it is impossible to extract hour, minute and second data, extract as much time data as you can.

Data Extraction and Input, Examples:

The figures below are meant to illustrate where to look for the information we require.

Figure 1. Identifying the video ID from a video URL.

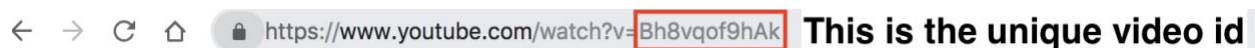


Figure 2. Identifying the video title and snippet.

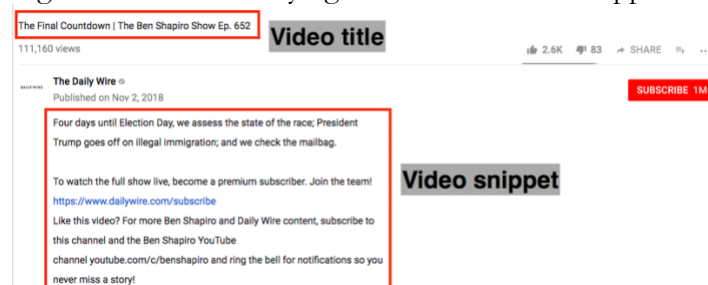


Figure 3. Identifying authors of comments and replies.

Cap - l'Élite

3 days ago

Comment op

Ben Shapiro is better at reviewing memes than Pewdiepie 🍌

👍 572

REPLY

Hide replies ^

Damon Andes

3 days ago

Reply op

Cap - l'Élite 🍌 meme review 🍌

👍 11

REPLY

I, Ben Shapiro. Want 10k subs without a vid.

3 days ago

Cap - l'Élite you know it

Figure 4. How the data looks on the spreadsheet

video_id	video_url	video_title	channel	series	video_snippet	video	comment	reply	video_op	comment_o	p	reply_op	date_posted	content	date_scrape
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		1	0	0	Ben Shapiro	NA	NA	2018_11_02	[1 minute transcript using 'transcribe'] Speaker 1: Four days until election day. We assess the state of the race President Trump goes off on an illegal immigration and we check the mail bag. I'm Ben Shapiro. This is the Ben Shapiro show. Speaker 2: So in honor of Halloween, we've decided to move our entire Studios Speaker 1: to the upside down. It's very exciting. We actually are rebuilding our entire studio will be massive and magnificent and we're doing all of it with your money which makes it even better. It's just spectacular we can get to all the news in just one second. But first let's talk about how you're going to relax this weekend. So this weekend this whole week and next week. It's gonna be pretty stressful right and that's why I occasionally you need to relax one of the best ways to relax is by getting a massage in your own home. Bring the spot to you. This is where Zeal comes in Zeal.com or Zeals iPhone or Android app Ze elle.com. It allows you to select from Top local licensed pre-screen massage therapist. Choose your favorite technique gender preference time and location for your massage Zeal will send one of their ten thousand licensed massage therapist with a massage table music supplies. They'll give you a five.	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	1	0	Ben Shapiro	Cap - l'Élite	NA		Ben Shapiro is better at reviewing memes than Pewdiepie 🍌	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	0	1	Ben Shapiro	Cap - l'Élite	Damon Andes		Cap - l'Élite 🍌 meme review 🍌	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	0	1	Ben Shapiro	Cap - l'Élite	I, Ben Shapiro. Want 10k subs without a vid.		Cap - l'Élite you know it	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	1	0	Ben Shapiro	fanaddicts of film	NA		Slowly increasing in epic quality.	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	0	1	Ben Shapiro	fanaddicts of film	SiWtard		He reached that peak long ago. What comes after epic?	
pRwJLQgQJlU	https://www.	The Final Countdown Ep. 652	The Daily Wire	The Ben Shapiro Show	Four days unt		0	0	1	Ben Shapiro	fanaddicts of film	Elliot Speavack		Transcendent @SiWtard	

Fields shaded grey correspond to the times the videos/comments/replies were posted/scraped. You can't extract those by copying and pasting from YouTube, but we expect that tuber or the API will allow you to access such metadata.