

# Automatic Sampling and Analysis of YouTube Data

## Tools for the Automatic Sampling of YouTube Data

Julian Kohne  
Johannes Breuer  
M. Rohangis Mohseni

2020-02-10

# Tools for the Automatic Sampling of YouTube Data

- Webometric Analyst
- YouTube Data Tools
- `tuber`

We wrote a [short tutorial](#) on how to use the tools. **NB:** The `tuber` script mentioned in that tutorial is a bit outdated. We will explore how to work with `tuber` in more detail in this workshop.

# Overview

Method	Manual Coding	Webometric Analyst	YouTube Data Tools	tuber
Type	n/a	Program	Web service	Package for R
Platforms	All	Win	All	Win, Mac, Linux, Unix
Collected Features	Depends on coding scheme	Channel Info, Video Info, Comments, Video Search	Channel Info, Video Info, Comments, Video List	Channel Info, Video Info, Comments, Subtitles, All searches
Scoping	Depends on coding scheme	100 most recent or all comments	All comments	20-100 most recent or all comments

# Pros and Cons

Method	Manual Coding	Webometric Analyst	YouTube Data Tools	tuber
Need API Key?	No	Yes	No	Yes
Disadvantages	Time-consuming	Only first 5 follow-up comments, no error feedback, undetectable time-outs	Lacking flexibility, fewer infos	Only first 5 follow-up comments due to bug
Ease of Use	High	Low	High	Low
License	n/a	Free for n/c	Open Source	Open Source
Example: Dayum Video (22-02-2019, 2pm)	47,163	44,828	47,153	44,810

Dayum Video / tuber bug

# YouTube Data Tools

## YouTube Data Tools

[blog](#) [software](#) [research](#) [DMI](#) [about](#)[Home](#) [Channel Info](#) [Channel Network](#) [Video List](#) [Video Network](#) [Video Info](#) [FAQ](#)

### Video Info and Comments Module

This module starts from a video id and retrieves basic info for the video in question and provides a number of analyses of the comment section. Comments are retrieved via the [commentThreads/list](#) API endpoint.

The number of comments the script is able to retrieve can vary wildly. In some cases, only a relatively small percentage is made available, while in others well over 100.000 comments have been successfully retrieved. This seems to be mainly related to the age of the video in question.

The module creates the following outputs:

- a tabular file containing basic info and statistics about the video;
- a tabular file containing all retrievable comments, both top level and replies;
- a tabular file containing comment authors and their comment count;
- a network file (gdf format) that maps interactions between users in the comment section;

The first three elements can be shown directly in the browser by enabling HTML output.

### Parameters

Video id:  (video ids can be found in URLs, e.g. <https://www.youtube.com/watch?v=aXnaHh40xnM>)

HTML output: ☐ (displays HTML result tables in addition to file exports)



Ich bin kein Roboter.



reCAPTCHA

[Datenschutzerklärung](#) - [Nutzungsbedingungen](#)

# How to Retrieve Comments with YouTube Data Tools

- Identify relevant videos
- Launch Video Info and Comments Module
- Insert relevant Video ID (one at a time, e.g. XWxRrCXjmF0)
- Save "...\_comments.tab" and rename it to "...comments.txt"
- Open it with Notepad, copy&paste to Excel and save as .XLS
- **NB:** Saving as .CSV destroys emojis

**Any questions?**