

I Stayed Up Until 3 AM To Refute The CDC

And Destroyed Their Arguments With A Bot In The Process

The Underdog Sep 22

12 16

If you've been wondering where I've been the last few days I've been quite busy.

Some many months back, I had decided to pull down and parse through the 1000+ studies that Dr. Paul Alexander had published, claiming they indicated COVID-19 shot harms. I wanted to verify study titles, ensure there weren't any duplicates, and make sure there weren't any irrelevant, unavailable, or withdrawn studies.

Now, this was a mammoth task, but I am no stranger to Big Data in Healthcare. I spent many months chipping away at the task. I converted the list into a spreadsheet, fixed broken links, removed unnecessary meta-data from URLs, parsed out duplicate links, and more. I reduced the number down to 898 studies but the parsing wasn't complete.

Then I made an insane attempt to manually review the studies - as seen in "Assessing 898 COVID-19 Shot Studies: Studies #1 to #11" and "Assessing 898 COVID-19 Shot Studies: Studies #12 to #22" - however, with the combination of real life and the demands of high quality Substack articles consuming free time, this wasn't realistic. It required more time and effort than I could give it.

That and people aren't big fans of dry, boring, numbered titles.

Enter The Bot

The main reason for reviewing the studies was to find and remove any duplicate papers that were available on multiple journals. With such a large dataset the odds of this occurring were quite high. Manually extracting the titles of nearly 900 studies is a mammoth task. Even if I could process 1 title a second, it'd take almost 15 minutes.

I decided I'd use my programming skills to write a Python script to use Selenium in order to webscrape the titles of the papers from the links. Using the assistance of GPT-3 to generate some CSV-JSON conversion scripts, I worked up until 10pm to get the bot running.

```
File Edit Search View Document Project Build Tools Help
Symbols Documents NIHAnalysis.py
Functions 225 return None
  Demostudies [274] 226
  FilterDuplicates [258] 227 def ProcessStudies(filename):
  GetLinkTitle [8] 228 TempJSON = ReadJSONFile(filename)
  ProcessStudies [227] 229 TB = PrepareBrowser("https://www.ncbi.nlm.nih.gov/",60)
  SearchCollectorForStudy 230 WriteFile = False
  Variables 231
  CSVFilename [291] 232 try:
  Filename [289] 233 for EachStudy in TempJSON:
  JSONFilename [290] 234 if len(EachStudy['Study Name']) == 0:
  235 TempLink = EachStudy['Study Link']
  236 LinkTitle = GetLinkTitle(TB,TempLink)
  237
  238 if LinkTitle is not None:
  239 EachStudy['Study Name'] = LinkTitle.strip()
  240 if not WriteFile:
  241 WriteFile = True
  242
  243 except Exception as err:
  244 print(err)
  245
  246 if WriteFile:
  247 WriteJSONFile(filename,TempJSON)
  248
  249 def SearchCollectorForStudyName(TempJSON,StudyName):
  250
  251
```

My original goal was to simply get as many titles as possible automatically, update the spreadsheet, then via a combination of automatic comparison and manual review, remove the duplicates, note the changes, publish the spreadsheet.

Essentially, a once-time job.

Or so I thought.

Shocking To Watch In Person

I kept tabs on what the bot did with the browser to see how well it was handling the studies. I squashed many bugs in the code - namely the bot complaining it didn't recognise a website and thus couldn't process it, requiring I add more rules.

Once it got running in full swing, it was a shocking horror show of studies to watch. Hundreds of studies - each one representing many months of work between many researchers and victims - flashed on my screen, and did not stop.

I sat there for over five minutes unable to look away, like watching a slow motion car crash. Every second a new study. The enormity of it all only dawning through the sheer length. I knew hundreds of studies was big, but it is just a number, it doesn't give a sense of scope.

This Was The Best Way To Present The Data

I had wondered what the best way was to present all this data in order to expose the various health agencies corruption. Watching the bot go through the studies struck me as the most effective given the enormity and length of it all.

However I realised I was running against the clock, as such activity was unlikely to go unnoticed and websites tend not to like aggressive browsing behaviour from a bot - even if it is in the interests of the public's health. I was going to have to pull an all-nighter.

I manually filled in the studies the bot was unable to parse, removed any broken text, skimmed the list to remove unnecessary HTML tags dragged in, and sorted by ascending order the titles to see if there were duplicates. There were, so I wrote more code for the bot to filter them out.

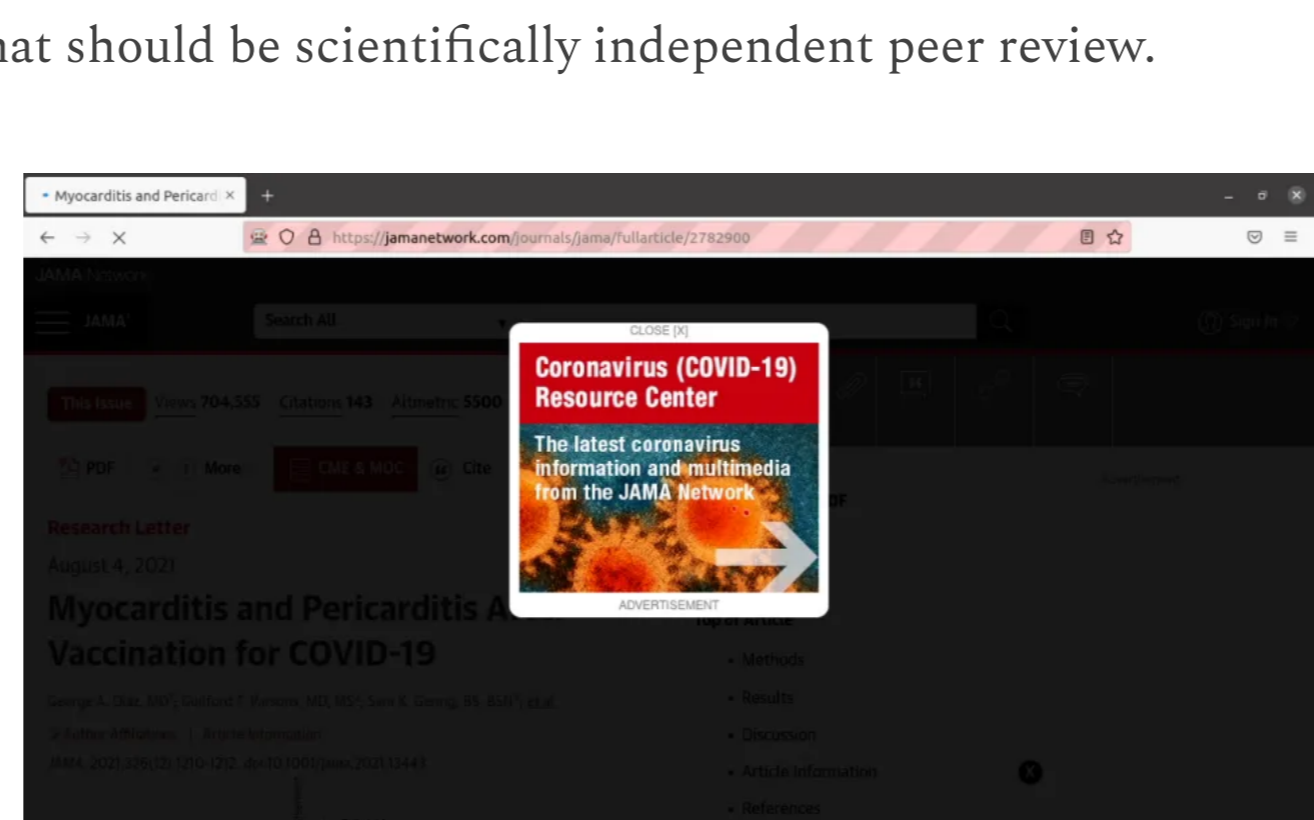
```
DuplicateStudiesRemoved.txt 34.62 KB Edit Lock Replace D
1 Metadata On Corrections:
2
3 755 study links remain
4 12 link typos corrected (spaces, missing hypens, incomplete URLs)
5 6 URLs updated
6
7 259 links removed, of which were:
8 2 Google documents (both censored by Google, no archived copies, not peer-reviewed)
9 1 irrelevant article relating to radionuclides (appears to be a duplicate link typo)
10 2 irrelevant articles relating to facemasks
11 4 withdrawn
12 84 duplicates removed by automated duplicate line removing tool
13 131 duplicates removed by advanced title grabbing and comparison tool
14 7 duplicates removed by human review
15 28 duplicate links obscured by metadata tags in URL or fullstops requiring human intervention
16
17
```

I manually updated my Gitlab text, reporting the changes, removals, updates, URL fixes, normalisations and more. I opened similar title studies to check authorship to ensure they were authored by different people and not duplicates. I had the bot rapidly flash the dwindling list of entries in webbrowser for manual review.

This easily took me up to midnight as I kept checking back and forth, checking the bot's handiwork and re-checking my own. One hour passed, then another. I whittled it down, from 898, to 756. The bot found 131 duplicates. I found 11 issues.

I Set Up The First Recording Run

Eager to avoid being blocked by websites for doing the right thing, I moved quickly to get something - anything - recorded on video. I set up a test recording run. In doing so I spotted another irrelevant study and purged that. Websites were clunky - so many adverts on what should be scientifically independent peer review.



JAMA Network were one of the biggest culprits

I sat for 15 minutes watching the dry run occur. Some sites loaded slowly. Some were blocked by cookie prompts. Others were broken. Bot edits, changes. I stuffed another snack to keep myself going. Spicy noodles. Acid reflux.

I re-wrote portions of the bot code, set up rules to skip specific parsing actions, dropped some websites (which had cottoned on) entirely, focused on the faster loading ones - mainly the NIH pages - instead. I updated the dataset so it was now 755 studies. It was now 3 AM and I ran the bot and video recording software again.

With Or Without Music?

So many injured people in so many studies just flickered by. Horrifying, but I couldn't waste time re-watching what I would see again during editing. Watching it without any sound was incredibly dry.

During this time I opened up my second, much, much older, laptop to look for suitable music for the video. Finding public domain or open source music that was longer than 7 minutes was difficult, especially fitting music.

I had very few options and listened to each one whilst watching the bot work. I really wanted to use a remix of "Popcorn" by Kraftwerk (I also quite liked Enzo's Oxylene remix) but being copyrighted - like many alternate renditions - it was not an option.

I also wasn't sure if the music was too upbeat for such morbid things. I had considered the symbolic "Killers" but it was only 5 minutes long. I ultimately settled on the public domain "Ambush in Rattlesnake Gulch", which being 10 minutes long and of a tempo matching the pace of viewed studies, seemed the most fitting.

Too Tired To Continue

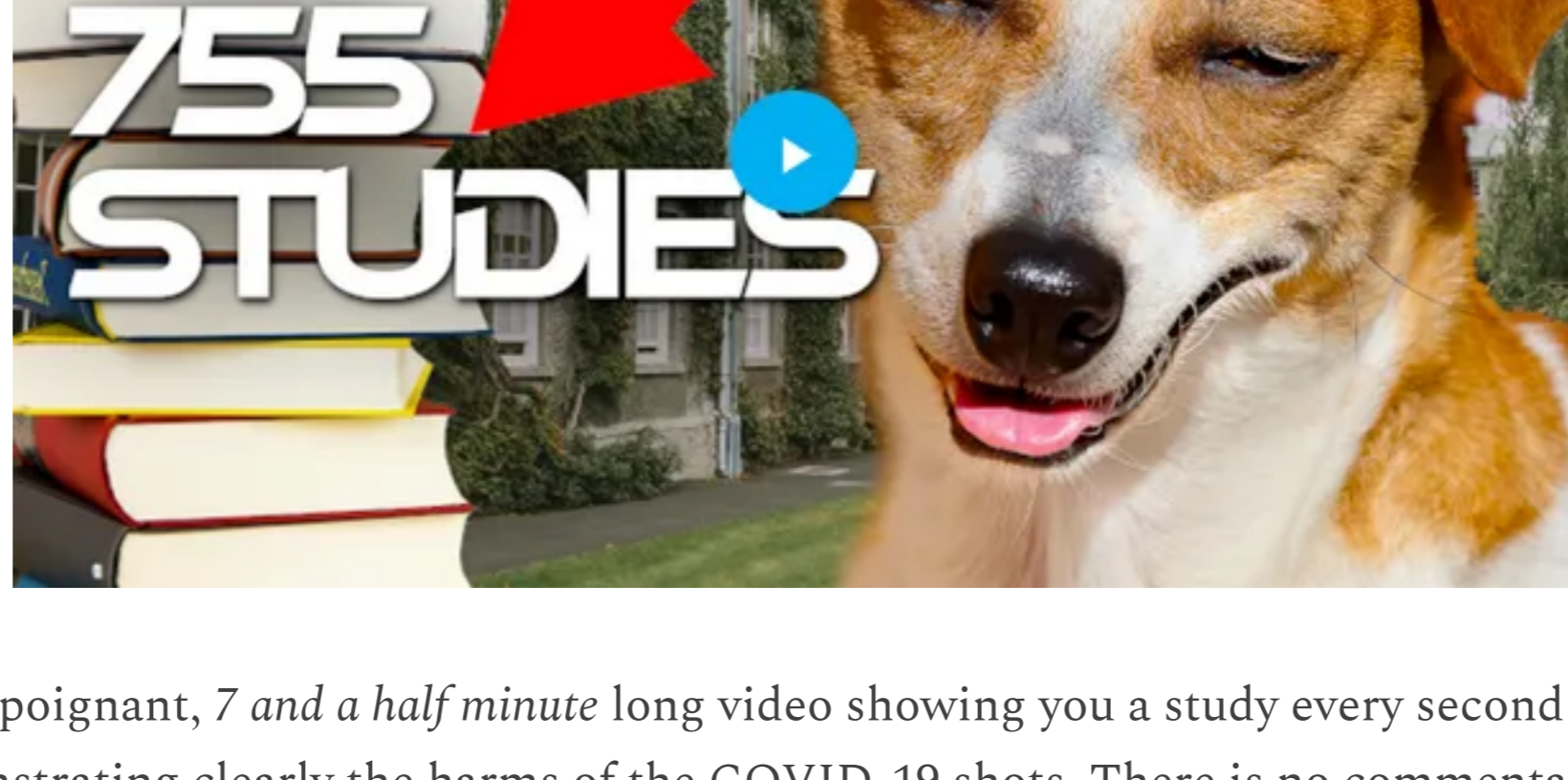
3 AM rolled around, and lack of sleep impairs judgement much like being drunk whilst driving does. I gotten the urgent video recording session I required before any corrupt cronies could adapt and prevent, so I could sleep, then edit and upload the video to Brighteon the next day to continue.

I threw it into the ever buggy and crashy kdenlive, and made an attempt at editing the clips together. Kdenlive corrupted a bunch which got tossed as a result, with mostly only the NIH studies surviving in the final outtake. At 7 minutes long it was lengthy enough to make a point.

I threw in "Ambush in Rattlesnake Gulch" and previewed the video to ensure no problematic errors were present, before uploading the video, whilst working on the thumbnail in GIMP, throwing together attention grabbing elements.

Now You Can Watch It - Share With Friends

You'll want to watch the video Refuting Health Agencies:



It is a poignant, 7 and a half minute long video showing you a study every second demonstrating clearly the harms of the COVID-19 shots. There is no commentary, only music so it isn't completely dry.

I hope the staying up to 3 am was worth it. Feel free to show to your friends - and more importantly, the health agencies themselves. The list of 755 studies showing harms with the COVID-19 shots can be found here.

Daily Beagle's is still a long way off from becoming financially viable, consider becoming a paying subscriber today!

derekcishop@hushmail.com Subscribe

Share this article and encourage others to subscribe!

Share

What did you think of the video? Got any other insights?

Leave a comment

12 likes 12 16

16 Comments

Brilliant work! Exceptional compilation. We are all underdogs these days. Thank you.

George Sep 23 · edited Sep 23 · Liked by The Underdog
Fantastic video!
I really like the music but an not familiar with it.
Please provide source.
Never mind. Just reread the article.
Getting a Ph.D lowered my reading skills.

Bacca400 · Writes Science News · Sep 22 · Liked by The Underdog
Wow! That's a lot of work!
> write a Python script to use Selenium in order to webscrape the titles of the papers from the links.
OMG. Programmer here. Isn't there any metadata in some webpages to get the title of the study, or some other data about the study?
I'd be willing to help you with programming tasks, I just use Perl. It's great at text processing. For example I can deduplicate studies as I've done deduplication of mailing addresses. What a task that was!

Can't you deduplicate by the DOI number or DOI link? Every study has one I believe and the DOI number is unique to a study. More info at https://doi.org
What about having several people make a growing spreadsheet of researchers (sometimes there are 10 authors), DOI number, DOI link, publish date, link to full study, link to abstract only (not all full studies are free), and 3-5 topics for each study. That's just off the top of my head without doing any analysis on the study data.

3 replies by The Underdog and others

14 more comments...

Top New Community

A Polite Call Out To Steve Kirsch [Kirsch Has Replied!]
Help me understand, these documents are important, right?
THE UNDERDOG · AUG 26 · 38 · 51 ·

Political Sleeper Cell Goes Live
It is all in the wording
THE UNDERDOG · SEP 20 · 13 · 2 ·

Have I Been Able To Redpill People?
I Respond To Steve Kirsch
THE UNDERDOG · SEP 13 · 17 · 4 ·

See all >

Ready for more?

derekcishop@hushmail Subscribe

Publish on Substack Get the app