

Old Dominion University

## ODU Digital Commons

---

Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics

NSF Research Experiences for Undergraduates (REU) Programs 2022

---

Summer 2022

### Did They Really Tweet That?

Caleb Bradford

*Old Dominion University*, [cbrad022@odu.edu](mailto:cbrad022@odu.edu)

Michael L. Nelson (Mentor)

*Old Dominion University*, [mnelson@odu.edu](mailto:mnelson@odu.edu)

Follow this and additional works at: [https://digitalcommons.odu.edu/reu2022\\_computerscience](https://digitalcommons.odu.edu/reu2022_computerscience)



Part of the [Databases and Information Systems Commons](#), and the [Social Media Commons](#)

---

#### Recommended Citation

Bradford, Caleb and Nelson, Michael L. (Mentor), "Did They Really Tweet That?" (2022). *Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics*. 1. [https://digitalcommons.odu.edu/reu2022\\_computerscience/1](https://digitalcommons.odu.edu/reu2022_computerscience/1)

This Presentation is brought to you for free and open access by the NSF Research Experiences for Undergraduates (REU) Programs 2022 at ODU Digital Commons. It has been accepted for inclusion in Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

# Did They Really Tweet That?



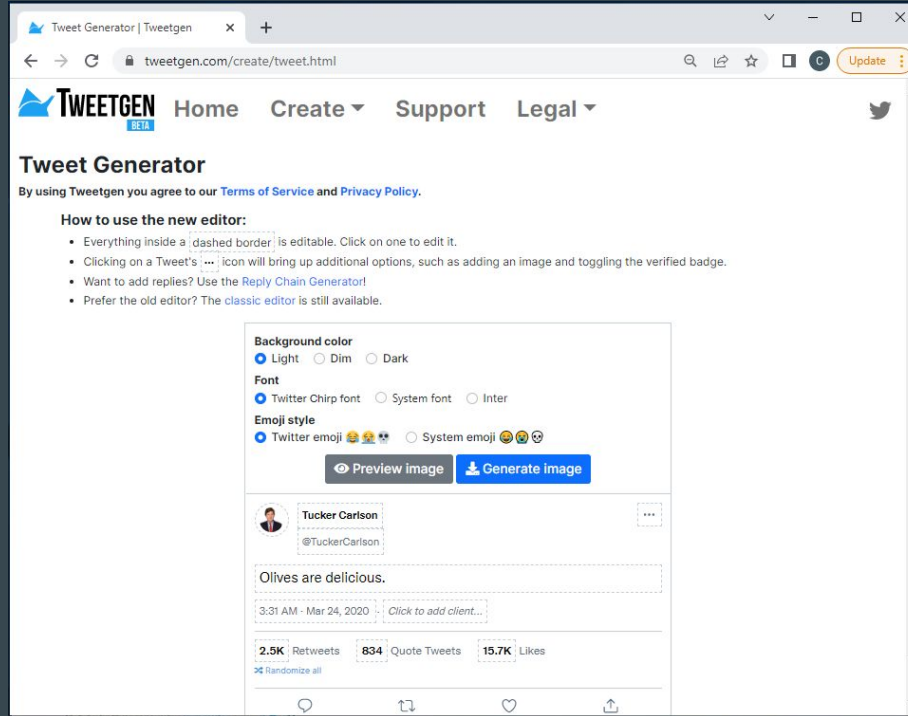
Caleb Bradford advised by Dr. Michael Nelson  
2022-08-04



# An interesting and totally real tweet made by a colleague

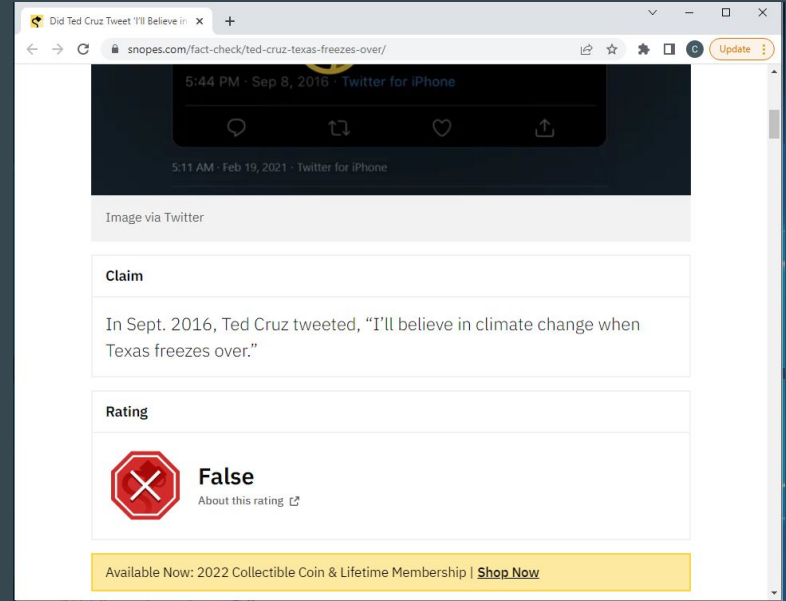


# Fabricated with easy to use web app called Tweetgen



<https://www.tweetgen.com/>

# Goal: Validating attribution



<https://www.snopes.com/fact-check/ted-cruz-texas-freezes-over/>

Given a screenshot of a tweet, we want to make sure that tweet was made by that person

# Easily spread by credible people



A now [deleted tweet](#) propagating the fake

# Twitter screenshots are prevalent across all social media



Twitter: QTING would provide engagement

Instagram: Sharing posts across platforms

# How can we verify attribution?

The image shows a screenshot of a tweet from Haley Bragg (@haleybragg17) with several red boxes highlighting key metadata. Arrows on the left point from labels to these boxes:

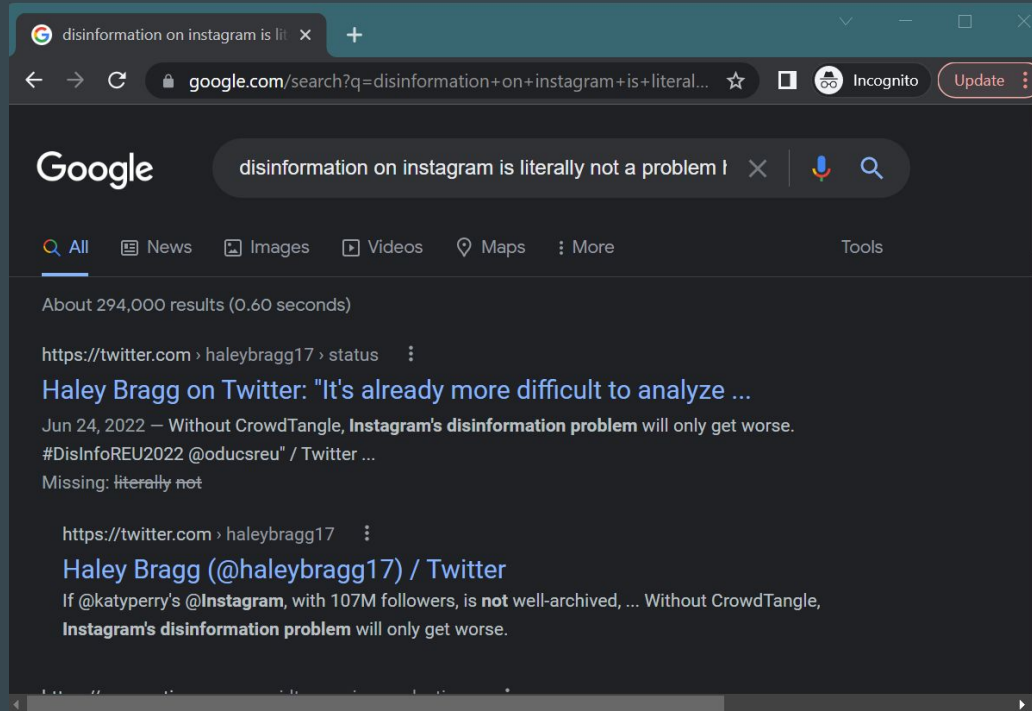
- Twitter handle:** Points to the user's name and handle, **Haley Bragg @haleybragg17**.
- Tweet body:** Points to the text of the tweet: **Disinformation on Instagram is literally not a problem.**
- Date and time:** Points to the timestamp: **11:25 AM · Jul 31, 2022**.
- Engagement:** Points to the interaction counts: **310 Retweets 214 Quote Tweets 4.6K Likes**.

At the bottom of the tweet are icons for replying, retweeting, liking, and sharing.

A lot of important metadata can be found in the screenshot itself



# There is a good chance the answer is on the Web



A common first instinct is to simply search Google for more information

# Some tweets supposedly made by well-known accounts attract the attention of fact-checkers

## Claim

U.S. President Donald Trump tweeted in November 2009 that "Obama's handling of this whole pandemic has been terrible!" and that he "would never let thousands of Americans die from a pandemic while in office."

## Rating



**False**

About this rating [↗](#)

## False claim: In 2009 Trump tweeted that he "would never let thousands of Americans die from a pandemic"

By Reuters Staff

3 MIN READ



Social media users are circulating a screenshot of what appears to be an old tweet from President Donald Trump ( [here](#) , [here](#) , and [here](#) ).

The image shows a tweet allegedly posted by Trump on November 23, 2009, during the swine flu outbreak. It reads: "Obama's handling of this whole pandemic has been terrible! As President, ALL responsibility becomes yours during a crisis like this, whether or not you're entirely to blame. John McCain, and for that matter myself, would never let thousands of Americans die from a pandemic while in office."

<https://www.snopes.com/fact-check/2009-trump-tweet-pandemic/>

Fact check from Snopes.com

<https://www.reuters.com/article/idUSKCN2242AK>

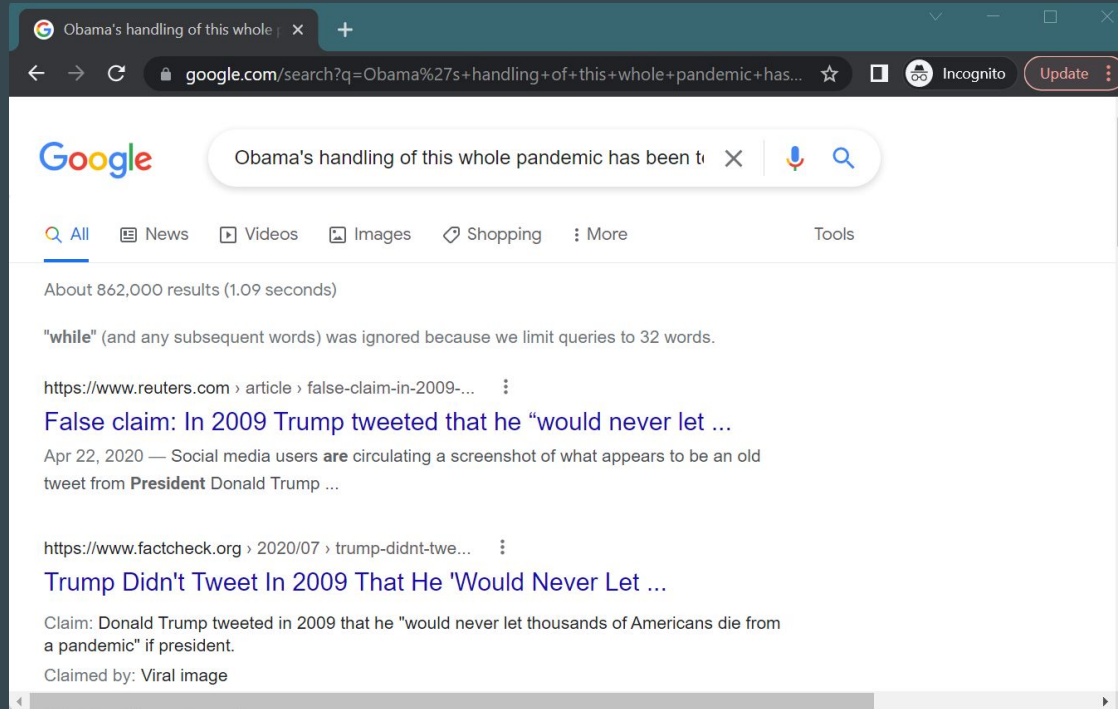
Fact check from Reuters.com



Using just the text of the alleged tweet, can we determine its veracity by querying search engines?




# A Google query using just the body of a tweet



[Live web version](#) [Archived version](#)

# Queries can also be made to sites with built-in search functions

Obama's handling of this whole pandemic has been terrible! As President, ALL responsibility becomes yours du 

## Search Results for "Obama's handling of this whole pandemic has been terrible! As President, ALL responsibility becomes yours during a crisis"

[NEWS](#)   [COMPANIES](#)   [VIDEOS](#)   [PICTURES](#)

1 Results   Sort by: **Relevance** ▾   Date Range: **Anytime** ▾

**False claim: In 2009 Trump tweeted that he "would never let thousands of Americans die from a pandemic"**

... pandemic has been terrible! As President, ALL responsibility becomes... the swine flu outbreak.  
It reads: "Obama's handling of this whole..."

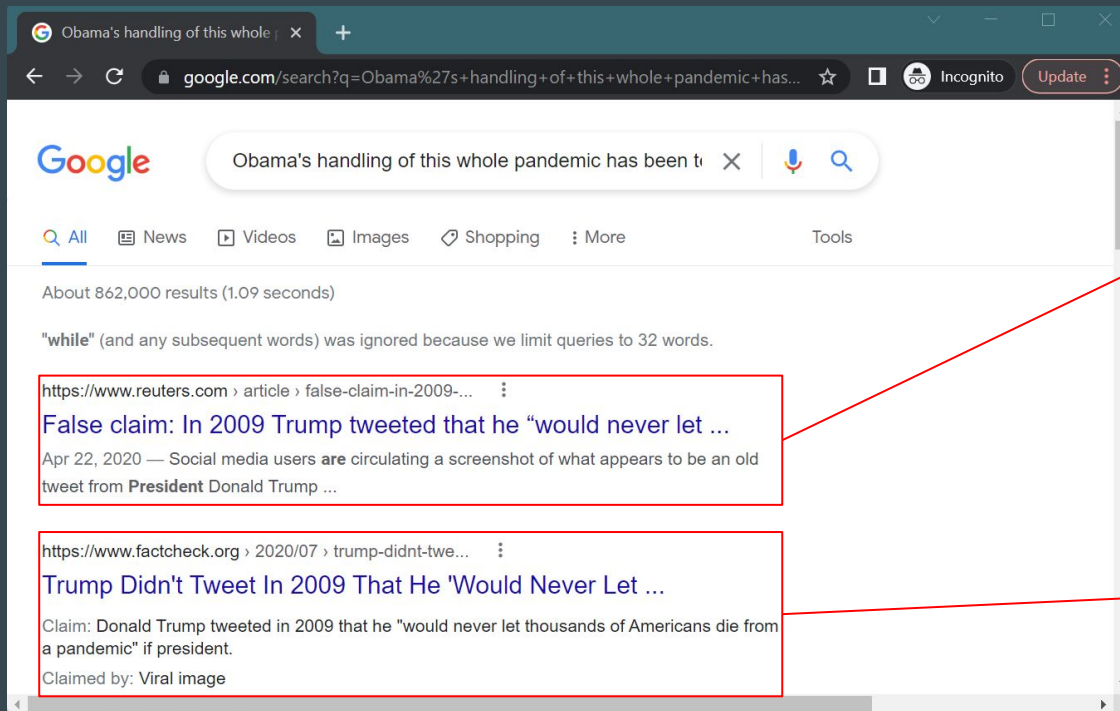
APRIL 22, 2020 10:51AM EDT

[Search query to Reuters](#)

[Archived version here](#)



# Important metrics in the efficacy of a search query



Mean reciprocal rank = 1  
Precision at 1 = 1

Mean reciprocal rank = 0.5  
Precision at 1 = 0

# Building a dataset of Snopes articles and search engine results

```
"Article 1":{
  "URL": "https://www.snopes.com/fact-check/donald-trump-fake-tweet/",
  "Supposed author": "Donald Trump",
  "Tweet body": "My grandparents didn't come to America all the way from",
  "Snopes query": {
    "Query URL": "https://www.snopes.com/?s=My+grandparents+didn%27t+co",
    "Reciprocal rank": 0,
    "Precision at 1": 0
  },
  "Google query (site:snopes.com)":{
    "Query URL": "https://www.google.com/search?q=My+grandparents+didn%",
    "Reciprocal rank": 1,
    "Precision at 1": 1
  },
  "Google query": {
    "Query URL": "https://www.google.com/search?q=My+grandparents+didn%",
    "Reciprocal rank": 1,
    "Precision at 1": 1
  },
  "Reuters query": {
    "Query URL": "https://reuters.com/search/news?blob=My+grandparents+",
    "Reciprocal rank": 0,
    "Precision at 1": 0
  }
}
```

[Full dataset available here](#)

Basic analysis of 30 data points found the following:

- Snopes Search
  - Mean reciprocal rank = 0.5500
  - Average P@1 = 0.5333
- Google query
  - Mean reciprocal rank = 0.4513
  - Average P@1 = 0.2667
- Google query with site operator
  - Mean reciprocal rank = 0.8667
  - Average P@1 = 0.8667
- Search on Reuters for related article
  - Mean reciprocal rank = 0.2167
  - Average P@1 = 0.2000



How can we automate the process of making queries given the text of an alleged tweet?





# Web queries to fact-checking sites can easily be made programmatically

Obama's handling of this whole pandemic has been terrible! As President, ALL responsibility becomes yours du

Search Results for "Obama's handling of this whole pandemic has been terrible! As President, ALL responsibility becomes yours during a crisis"

NEWS COMPANIES VIDEOS PICTURES

1 Results Sort by: Relevance Date Range: Anytime

**False claim: In 2009 Trump tweeted that he "would never let thousands of Americans die from a pandemic"**

... pandemic has been terrible! As President, ALL responsibility becomes... the swine flu outbreak.

It reads: "Obama's handling of this whole..."

APRIL 22, 2020 10:51AM EDT

```
Command Prompt
C:\Users\caleb\OneDrive\Documents\GitHub\REU-Files\ToolsForReuters>python3 query_reuters.py InputFiles/tweet.txt
Query being made: https://reuters.com/search/news?blob=Obama%27s+handling+of+this+whole+pandemic+has+been+terrible%21+As+President%2C+ALL+responsibility+becomes+yours+during+a+crisis
Article found at URL: http://reuters.com/article/idUSKCN2242AK
C:\Users\caleb\OneDrive\Documents\GitHub\REU-Files\ToolsForReuters>
```

The same query to Reuters being made via Python script (tweet passed from input file CL argument)



# HTML of articles can be scraped for their “truth rating”

## Claim

U.S. President Donald Trump tweeted in November 2009 that "Obama's handling of this whole pandemic has been terrible!" and that he "would never let thousands of Americans die from a pandemic while in office."

## Rating



**False**

About this rating [↗](#)

```
Command Prompt
C:\Users\caleb\OneDrive\Documents\GitHub\REU-Files\ToolsForSnopes>python3 query_snopes.py InputFiles/tweet.txt
Article found at URL: https://www.snopes.com/fact-check/2009-trump-tweet-pandemic/
Truth rating: False

C:\Users\caleb\OneDrive\Documents\GitHub\REU-Files\ToolsForSnopes>
```

```
<div class="media position-relative"> flex
  <div class="avatar-md avatar-lg-large mr-3"> == $0
    <figure class="figure w-100 mb-0">
      
    </figure>
  </div>
```

Python script scrapes article for rating

# Google does not appreciate non-humans making queries

```
cs_cbrad022@ganymede: ~
cs_cbrad022@ganymede:~$ curl https://www.google.com/search?q=this+is+a+google+search
<DOCTYPE html><html lang=en><meta charset=utf-8><meta name=viewport content="initial-scale=1, minimum-scale=1, width=device-width"><title>Error 403 (Forbidden)!!1</title><
style>{*{margin:0;padding:0}html,code{font:15px/22px arial,sans-serif}html{background:#fff;color:#222;padding:15px}body{margin:7% auto 0;max-width:390px;min-height:180px;pad
ding:30px 0 15px}* > body{background:url(//www.google.com/images/errors/robot.png) 100% 5px no-repeat;padding-right:205px}p{margin:11px 0 22px;overflow:hidden}ins{color:#77
7;text-decoration:none}a img{border:0}@media screen and (max-width:772px){body{background:none;margin-top:0;max-width:none;padding-right:0}#logo{background:url(//www.googl
e.com/images/branding/googlelogo/1x/googlelogo_color_150x54dp.png) no-repeat;margin-left:-5px}@media only screen and (min-resolution:192dpi){#logo{background:url(//www.goog
le.com/images/branding/googlelogo/2x/googlelogo_color_150x54dp.png) no-repeat 0% 0%/100% 100%;-moz-border-image:url(//www.google.com/images/branding/googlelogo/2x/googlelog
o_color_150x54dp.png) 0}}@media only screen and (-webkit-min-device-pixel-ratio:2){#logo{background:url(//www.google.com/images/branding/googlelogo/2x/googlelogo_color_150x
54dp.png) no-repeat;-webkit-background-size:100% 100%}}#logo{display:inline-block;height:54px;width:150px}</style><a href=//www.google.com/><span id=logo aria-label=Google>
</span></a><p><b>403.</b> <ins>That's an error.</ins><p>Your client does not have permission to get URL <code>/search?q=this+is+a+google+search</code> from this server. (C
lient IP address: 128.82.10.239)<br><br>
Please see Google's Terms of Service posted at https://policies.google.com/terms
<BR><BR><P>If you believe that you have received this response in error, please <A HREF="https://www.google.com/support/contact/user?hl=en">report</A> your problem. However
, please make sure to take a look at our Terms of Service (http://www.google.com/terms_of_service.html). In your email, please send us the <b>entire</b> code displayed belo
w. Please also send us any information you may know about how you are performing your Google searches-- for example, "I'm using the Opera browser on Linux to do searches f
rom home. My Internet access is through a dial-up account I have with the FooCorp ISP." or "I'm using the Konqueror browser on Linux to search from my job at myFoo.com. M
y machine's IP address is 10.20.30.40, but all of myFoo's web traffic goes through some kind of proxy server whose IP address is 10.11.12.13." (If you don't know any infor
mation like this, that's OK. But this kind of information can help us track down problems, so please tell us what you can.)</P><P>We will use all this information to diagn
ose the problem, and we'll hopefully have you back up and searching with Google again quickly!</P>
<P>Please note that although we read all the email we receive, we are not always able to send a personal response to each and every email. So don't despair if you don't he
ar back from us!</P>
<P>Also note that if you do not send us the <b>entire</b> code below, <i>we will not be able to help you</i>.</P><P>Best wishes,<BR>The Google Team<BR></P><BLOCKQUOTE>/+
/
```

## A curl request to a Google query that returns an error page



# Making Google queries is possible using software developed by Dr. Alexander Nwala

```
cs_cbrad022@ganymede:~$ scraper google "This is a google query"

googleSearchMain():

googleGetSERPResults() - start
  page: 1
  searchQuery: https://www.google.com/search?as_q=This%20is%20a%20google%20query
  derefURI(), sleep: 1

googleGetSERPResults() - end

printPayload(), src: google
1 https://support.google.com/docs/answer/3093343?hl=en
2 https://www.benlcollins.com/spreadsheets/google-sheets-query-sql/
3 https://www.youtube.com/watch?v=oCKvAcXTLZo
4 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=0
5 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=43
6 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=117
7 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=281
8 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=507
9 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=619
10 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=666
11 https://www.youtube.com/watch?v=oCKvAcXTLZo&t=692
12 https://www.youtube.com/watch?v=bW6P2YvLyZg
13 https://www.youtube.com/watch?v=W8C9TcYt47M
14 https://codingisforlosers.com/google-sheets-query-function/
15 https://codingisforlosers.com/google-sheets-query-function/#1_The_basics_of_Google_Sheets_Queries
16 https://codingisforlosers.com/google-sheets-query-function/#6_Building_Pivot_Tables_with_Query
17 https://www.howtogeek.com/450465/how-to-use-the-query-function-in-google-sheets/
18 https://spreadsheetpoint.com/google-sheets-query-function/
19 https://developers.google.com/chart/interactive/docs/querylanguage
20 https://sheetsformarketers.com/google-sheets-formulas/query/
21 https://coefficient.io/how-to-google-sheets-query-function/
22 https://cloud.google.com/bigquery/docs/samples/bigquery-query-external-sheets-perm
```

Dr. Nwala's "scraper" tool in action





# Using Nwala's work, broad searches for evidence becomes easier

```
cs_cbrad022@ganymede: ~/REUStuff/ToolsForGoogle
cs_cbrad022@ganymede:~/REUStuff/ToolsForGoogle$ cs_cbrad022@g
anyমেদে:~/REUStuff/ToolsForGoogle$ python3 query_google.py py
thon3 query_google.py InputFiles/tweet.txt
Reuters article found at URL: https://www.reuters.com/article
/uk-factcheck-trump-tweet-thousands-die/false-claim-in-2009-t
rump-tweeted-that-he-would-never-let-thousands-of-americans-d
ie-from-a-pandemic-idUSKCN2242AK
Verdict found in this article is False

Snopes article found at URL: https://www.snopes.com/fact-check
k/2009-trump-tweet-pandemic/
Truth rating in this article is False

cs_cbrad022@ganymede:~/REUStuff/ToolsForGoogle$
```

Google script can find variety of fact-check articles



# Future work

- Development of a more generic method of parsing fact-check articles for “truth rating”
  - Scraping raw HTML works, but needs to be specifically developed for every site we want to work with
  - The structure of a site is subject to constant change
- Building a more comprehensive dataset of known tweets
  - More tweets and their related queries
- Are there more effective ways to query the web for evidence?



# Conclusions

- Tweet misattribution is a unique form of disinformation that is extremely prevalent on all social media
- Queries built using just the body of a well covered tweet have shown efficacy at finding them
- Tools developed and further work available here
  - <https://github.com/calebkbrad/Attribution-Scraping>
- Acknowledgements
  - Dr Michael Nelson
  - Emily Escamilla
  - Tarannum Zaki
  - Dr Alexander Nwala
  - My fellow cohort members

