

OSINT Dashboard: Browser

Bryan Cruz Castillo
Cooper Van Kampen
Daniel Denekew
George Lonngren
Henry Post
Jarron Bailey
Michael Kotlyar
Rawad Alahmadi
Robert Bacias





Agenda

- Introduction
- What is Bowser
- Developer Workflow
- Web API for Programmers
- Web interface for everyone
- Data Sources
- Data Analysis
- Project Management Structure
- Communications
- Risks
- Possible Future Upgrades

What is Bowser?

Not a turtle....



Why the Name Bowser?



Mary "Bowser",
Union Spy
1846-1867



Executive Summary - What is Bowser?

It is an Open Source Intelligence (OSINT) Tool that searches controversial forums and boards for issues related to domestic Terrorism.



Really, what is Bowser?

- Bowser is a combination of a web API backend and an HTML frontend.
- It allows you to gather user data from different sources and analyze it for racism, terrorism, and national security issues
- We use keyword extraction to achieve this
 - This is a naive text analysis method
- Source code located at github.com/Team-Bowser-ITMS-448/



Try Bowser

Try it yourself! Go to:

tinyurl.com/browserweb for Web UI,

tinyurl.com/browserapi for API

What We Scraped

4chan.org and Reddit.com

4chan Message Boards: /pol/ and /x/ main focus

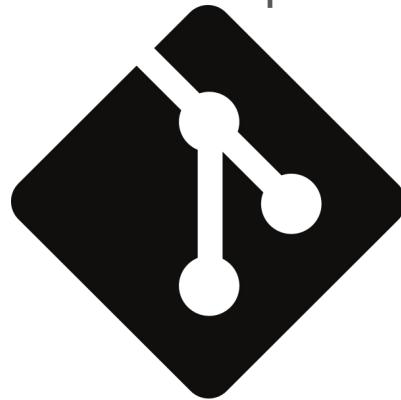
/pol/ - Politically Incorrect, Offensive and harmful content

/x/ - Paranormal, Conspiracy Theories



Developer workflow

- Git was used for source code control.
 - One main repository with branch protection and CI
- Fork and PR workflow used to ensure all developers can develop in their own style



Developer CI

Used Travis CI to run Python unit tests on code prior to allowing developers to incorporate code into main branch.

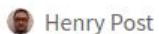
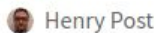


build passing

coverage 77%



Team-Bowser-ITMS-448 / ITMS448-osint-dashboard

build passingCurrent Branches Build History Pull RequestsMore options✓ masterHenry PostMerge pull request #71 from Team-Bowser-ITMS-→ #206 passed→ a564517🕒 1 min 39 sec📅 59 minutes ago✓ HenryFBP-patch-Henry PostUpdate README.md→ #204 passed→ b28a8b5🕒 1 min 38 sec📅 about an hour ago✓ masterJarron BaileyMerge pull request #70 from jarronb-iit/master→ #202 passed→ 2ed211a🕒 1 min 36 sec📅 2 hours ago✓ masterJarron BaileyMerge pull request #68 from jarronb-iit/master→ #200 passed→ c4925b4🕒 1 min 30 sec📅 4 hours ago✓ masterJarron BaileyMerge pull request #67 from jarronb-iit/master→ #197 passed→ 3098511🕒 1 min 35 sec📅 15 hours ago✓ masterHenry PostMerge pull request #66 from HenryFBP/master→ #195 passed→ 73385c4🕒 1 min 37 sec📅 a day ago

☐  **update to client** ✓

#70 by jarronb was merged 2 hours ago

☐  **added notification to gloabal state** ✓

#68 by jarronb was merged 4 hours ago

☐  **update to client** ✓

#67 by jarronb was merged 15 hours ago

☐  **add example url for reddit** ✓

#66 by HenryFBP was merged yesterday

☐  **conspiracy** ✓

#65 by kotlIT was merged yesterday

☐  **finish reddit json route** ✓

#64 by HenryFBP was merged yesterday

☐  **separate 4chan and reddit json api, add /api to routes** ✓

#63 by HenryFBP was merged yesterday



Web API for Developers

- We expose a web api that is self-documenting
- It can be tested at browser-web-app.freemyip.com:1839/api
- This serves CSV files and JSON files from Reddit and 4Chan
 - It embeds text analysis inside of the JSON files

+ - [View source](#)

```
{
  - example-urls: [
    "/api/generate/4chan/csv?
    boards=x,pol&flaggers=NSA_PRISM,TERRORISM&start_page=3&stop_page=10",
    "/api/generate/4chan/json?
    boards=pol,s4s,x&flaggers=NSA_ECHELON,RACISM&start_page=1&stop_page=3",
    "/api/generate/reddit/json?subreddit=Sino"
  ],
  message: "Welcome to the Bowser OSINT Web API! This is the index! See /routes/ for routes.",
  read-more-url: "https://github.com/Team-Bowser-ITMS-448/ITMS448-osint-dashboard/",
  route-url: "/api/routes"
}
```

```
{  
  desc: "This is a list of routes available to you to consume.",  
  - routes: [  
    "/api/generate/reddit/json",  
    "/api/generate/4chan/json",  
    "/api/generate/4chan/csv",  
    "/api/show/4chan/content-flaggers",  
    "/api/show/4chan/boards",  
    "/api/health",  
    "/api/routes",  
    "/api",  
    "/static/<path:filename>",  
    "/<path:path>",  
    "/"  
  ]  
}
```



```
    },  
    - full_comment: {  
      0: "This is pretty bad guys. \nhttps://m.youtube.com/watch?v=NGiJP9jGdUU",  
      1: "and it's beautiful",  
      2: null,  
      3: "https://www.youtube.com/watch?v=rqR1cjuPXUg\nthread theme",  
      4: "Do we know who the fuckfaces are that created and allowed this Marxist propaganda to be put out? http",  
      5: "https://wtvr.com/2019/12/02/1-in-5-us-adolescents-is-now-prediabetic-study-says-2/\nThat's a fu  
fructose corn syrup?\nMaybe my own country should stop buying food with that garbage in it too. \nSometh  
being fat but they're just ahead of the curve(sic)&#44; many western nations buy food from them or  
not an attack vector? Is this not making whites weaker? Who is profiting from this? And why does the US  
this situation. \nAnd again&#44; today it's America&#44; tomorrow it's the rest of us. \nThis  
ever saw anybody fat and when we did see a fatty&#44; it was strange and everybody noticed. \nNow I see  
We're being physically disabled by our very sustenance. We are being fed literal poison. \nFat is n  
that's a medical fact. \nWhat the fuck can we do? \nWe can't trust our governments to help sin  
government that socializes the expense of corn.",  
      6: "Kamala Harris drops out of the presidential race.\n\n>drops out before the Iowa Caucus \nBAHAHAHAHAHA  
CANDIDATE!\n\nhttps://www.foxnews.com/politics/kamala-harris-drops-out-of-presidential-race",  
      7: "How do we win more Hispanic voters&#44; especially Hispanic women?",  
      8: ">>235095610\nHispanic women are super cute and religious&#44; why do mutts want to deport them now a  
      9: ">>235095610\nYou don't you just deport and or kill them&#44; in minecraft of course.",
```


Web User Interface Demo



Web User Interface

- Built in ReactJS
- Utilizing:
 - Material UI
 - Easy-Peasy for Application State Management
 - React Chartkick for Data Charts
 - Formik for Form Management
 - Yup, for Object Schema Validation

Web interface - Form Configuration

Configure 4chan Data Query

Adjust settings

Host Configuration

Host

http://browser-web-app.freemyip.com

Port

1839

Generate CSV

Boards

☒ Advice

☒ Politics

☒ Television & Film

☒ Flash

☒ S*** 4Chan Says

☒ Paranormal

☒ High Resoulution

☒ Sports

☒ Auto

☒ Traditional Games

Flaggers

☒ NSA PRISM

☒ Hate Speech

☒ NSA ECHELON

☒ Racism

☒ Terrorism

☒ Conspiracy

Page Filters

Start Page

1

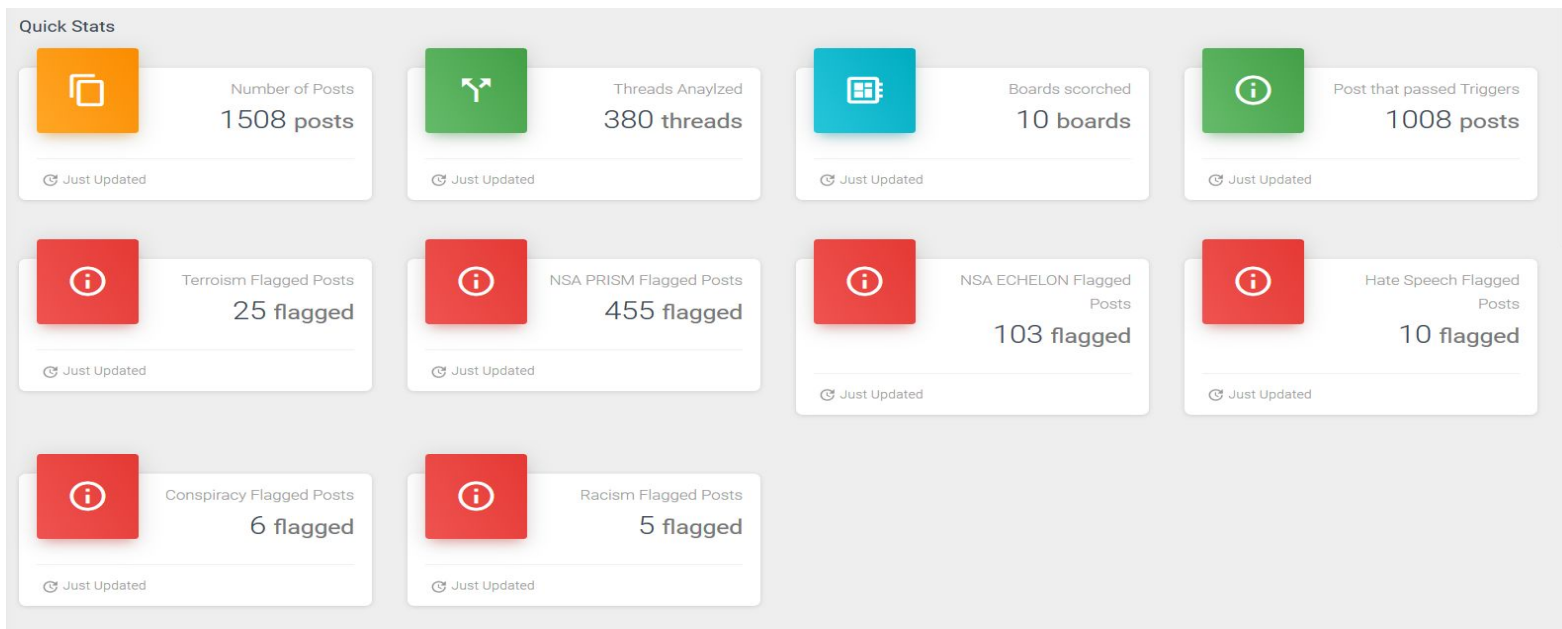
End Page

3

Request String: http://browser-web-app.freemyip.com:1839/api/generate/4chan/csv?boards=adv,pol,tv,f,s4s,x,hr,sp,o,tg&flaggers=NSA_PRISM,HATE_SPEECH,RACISM,NSA_ECHOLON,TERRORISM,CONSPIRACY&start_page=1&stop_page=3

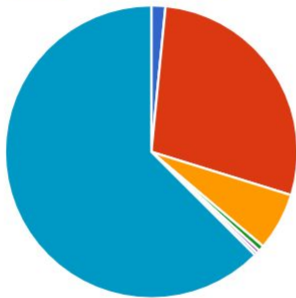
nin/dashboard

Web interface - Quick Stats



Web interface - Charts

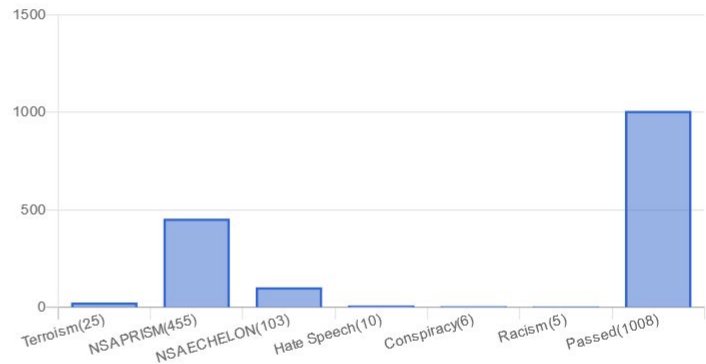
■ Terroism (25) ■ NSA PRISM(455) ■ NSA ECHELON(103)
■ Hate Speech(10) ■ Conspiracy(6) ■ Racism (5) ■ Passed(1008)



Pie Chart

Data from 4chan

🔄 Just Updated



Column Chart

Data from 4chan

🔄 Just Updated

Web interface - Data Table

Data Table

Posts Stats

POSTS

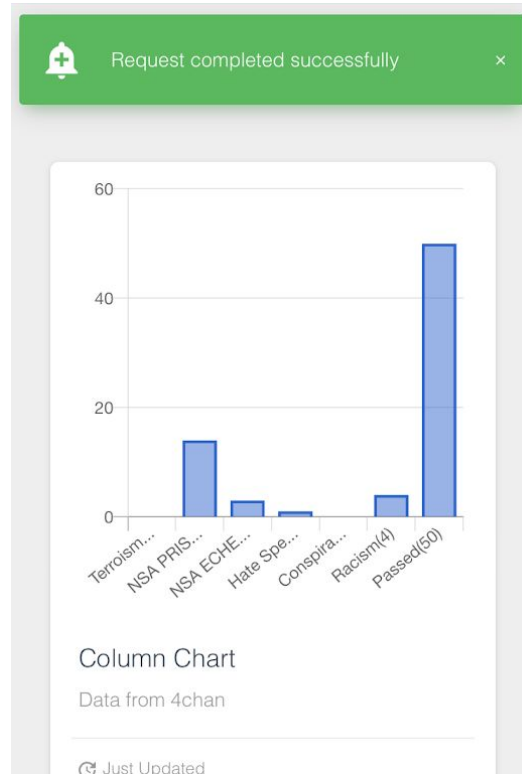
Search

Drag headers here to group by

	timestamp_epoch	timestamp_ISO8601	thread_url	thread_id	thread_api_url	post_url	post_id	post_api_url	op
>	1575413224	20191203T224704	http://archive.4plebs.org/adv/thread/21643186/	21643186	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21643186	http://archive.4plebs.org/adv/thread/21643186/#21643186	21643186	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21643186	True
>	1575413399	20191203T224959	http://archive.4plebs.org/adv/thread/21643186/	21643186	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21643186	http://archive.4plebs.org/adv/thread/21643186/#21643193	21643193	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21643193	False
>	1575412922	20191203T224202	http://archive.4plebs.org/adv/thread/21643166/	21643166	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21643166	http://archive.4plebs.org/adv/thread/21643166/#21643166	21643166	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21643166	True
>	1575413071	20191203T224431	http://archive.4plebs.org/adv/thread/21643166/	21643166	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21643166	http://archive.4plebs.org/adv/thread/21643166/#21643178	21643178	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21643178	False
>	1575412869	20191203T224109	http://archive.4plebs.org/adv/thread/21643165/	21643165	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21643165	http://archive.4plebs.org/adv/thread/21643165/#21643165	21643165	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21643165	True

5 rows | 1-5 of 1508

Mobile Friendly





Data Sources



- Reddit
 - Very popular
 - Has had controversial boards before
 - Globally used
- 4chan
 - At times it is used to share illegal and immoral material like child sexual abuse imagery
 - People regularly post about the gov't and terrorism
 - There have been incidents where 4chan posts have tipped authorities off to a crime





Recent events related to 4chan

- QAnon Conspiracy Theory
- 8chan - Controversial site created from banned 4chan topics
 - <https://slate.com/technology/2019/08/el-paso-8chan-4chan-mass-shootings-manifesto.html>
 - 8chan has been taken down from public internet (clearnet)

USERS

8chan Is a Normal Part of Mass Shootings Now

Three white nationalist shooters appear to have posted manifestos on the same troll forum. Here's how 8chan became a meme-filled refuge for hate.

By APRIL GLASER

AUG 04, 2019 • 1:11 AM

QAnon





Conspiracy's | QAnon | Pizzagate | *.gate

- [NBC News Story on QAnon](#)

How three conspiracy theorists took 'Q' and sparked Qanon

Pushing the theory on to bigger platforms proved to be the key to Qanon's spread – and the originators' financial gain.

- [New Conspiracy Theory Trending](#)

‘Epstein Didn’t Kill Himself’ and the Meme-ing of Conspiracy

What happens when a conspiracy theory becomes flattened into a quick, shareable phrase?

Second Source: Reddit

Currently Scraping /r/Sino

A Pro-China nationalist Subreddit for Westerners in the US/CA.

Users mentioning America get banned





Recent Events Related to /r/Sino

USA v China Trade War

Chinese Espionage in the US

[Petition on Change.Org to ban the subreddit](#)

Ban r/Sino on Reddit

r/Sino • 3m

You have been banned from participating in [r/Sino](#). You can still view and subscribe to [r/Sino](#), but you won't be able to post or comment.

Note from the moderators:

194 have signed. Let's get to 200!

First name
Last name

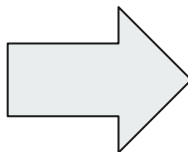
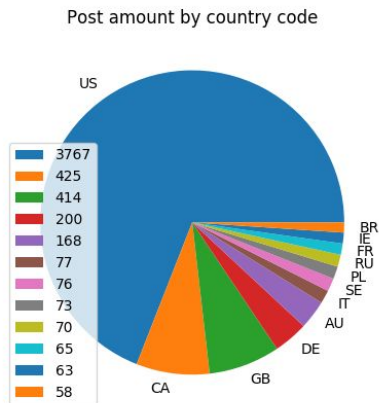


Data Analysis

- Most people are from the US who go on 4chan
 - We are not sure if people use VPNs to route through US
 - We cannot keep track of individual users (No ID)

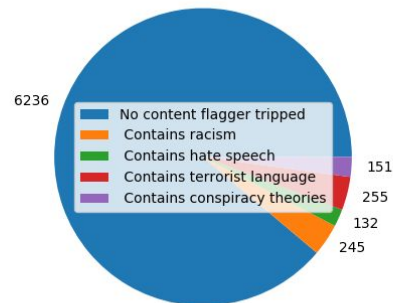
Data Analysis Example

Categorize Data by Location



Category breakdown by location

Detection breakdown by category, with benign posts



Data Analysis Errors

- **False Positives (Type 1 Error):** Null hypothesis is true, but rejected
 - Few instances during data parsing that the system returns a false negative.
 - Type 1 triggers: Users that utilize flagged keywords in reference to themselves. Keywords can hold a different connotations.

timestamp_epoch	timestamp_ISO8601	thread_url	thread_id	thread_api_url	post_url	post_id	post_api_url	op	country_code	board	[content flagger] Contains racism	[content flagger] Contains hate speech
✓ 1575487194	20191204T191954	http://archive.4plebs.org/adv/thread/21645760/	21645760	http://archive.4plebs.org/_api/chan/thread/?board=adv&num=21645760	http://archive.4plebs.org/adv/thread/21645760/#21645760	21645760	http://archive.4plebs.org/_api/chan/post/?board=adv&num=21645760	True		adv	False	True

Comment: My high school girlfriend updated her linkedin and it brought back all the high school memories and I just want to coom buckets on her but we broke up in high school and it has been 10 years since. We haven't talked but we are "friends" on social media. Is there a way to approach without looking like a stalking faggot or should I continue to coom solo?

Go to: <http://archive.4plebs.org/adv/thread/21645760/#21645760>



Data Analysis Errors (Cont.)

- **False Negatives (Type 2 Error):** Null hypothesis is false, but fails to be rejected
 - Frequency cannot be determined. We have to physically read all thread comments to find missed flags.
 - Type 2 triggers: Misspelling of keywords flags, new slang/terms not in library, words utilized by small knit communities.

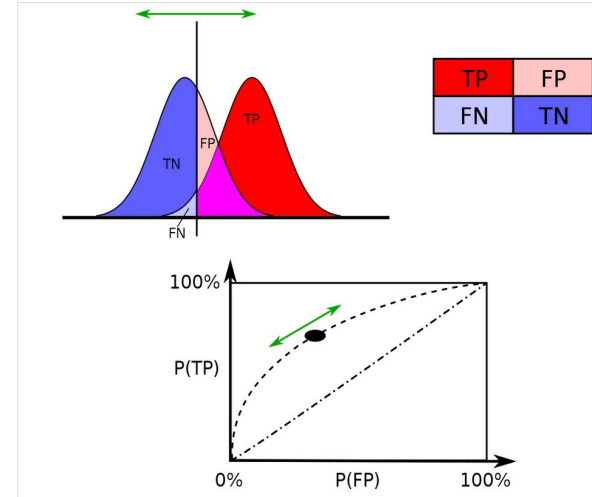


Data Analysis Error Future Fix Implementation

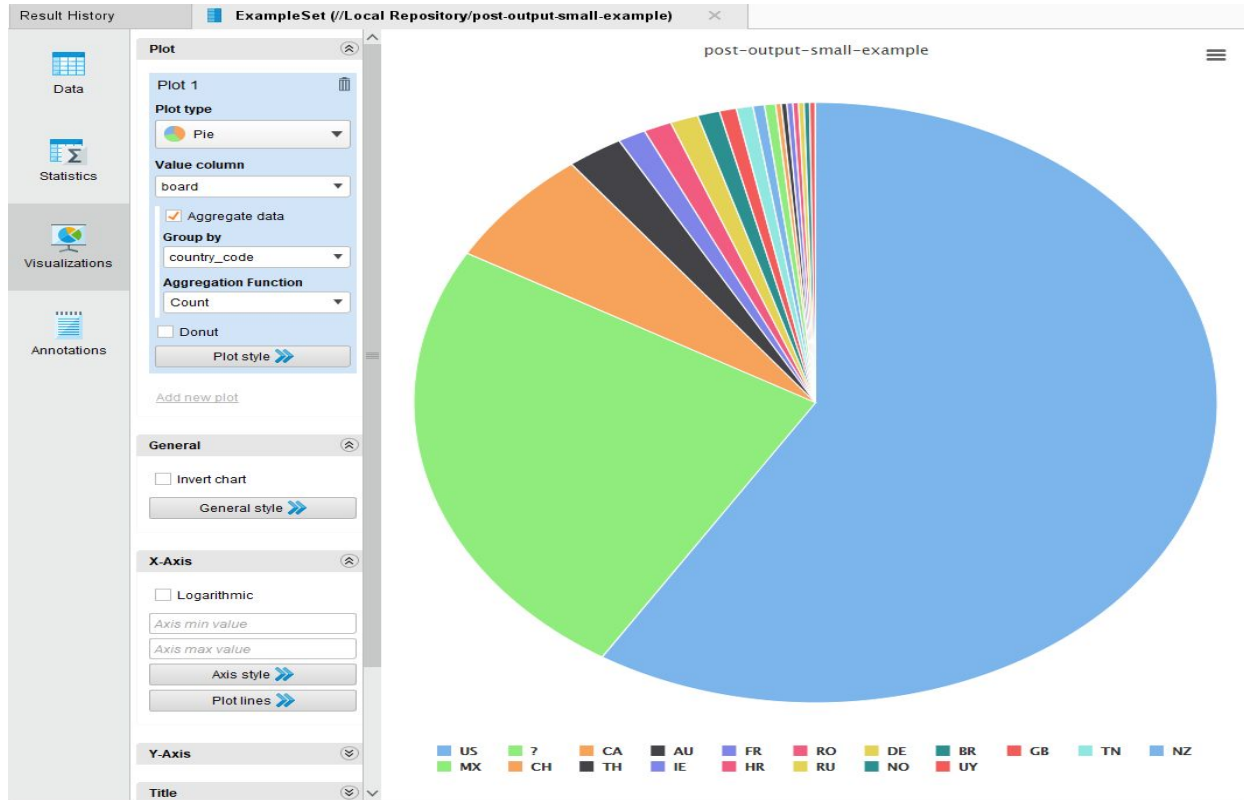
- Type 1 Errors:
 - Ignore strings of words that contain self reference
- Type 2 Errors:
 - Update Flagged Keywords Dictionary with new terms
 - Keep a library of dictionary terms with different variations they may appear as.
 - Ex.) Usage of symbols and numbers in flagged words

Data Analysis Errors

- We rather have more False Negatives than False Positives
 - Minimize the amount of “garbage” data being returned
- A threshold can make the data more specific or more sensitive, but risk factor will decrease/increases.



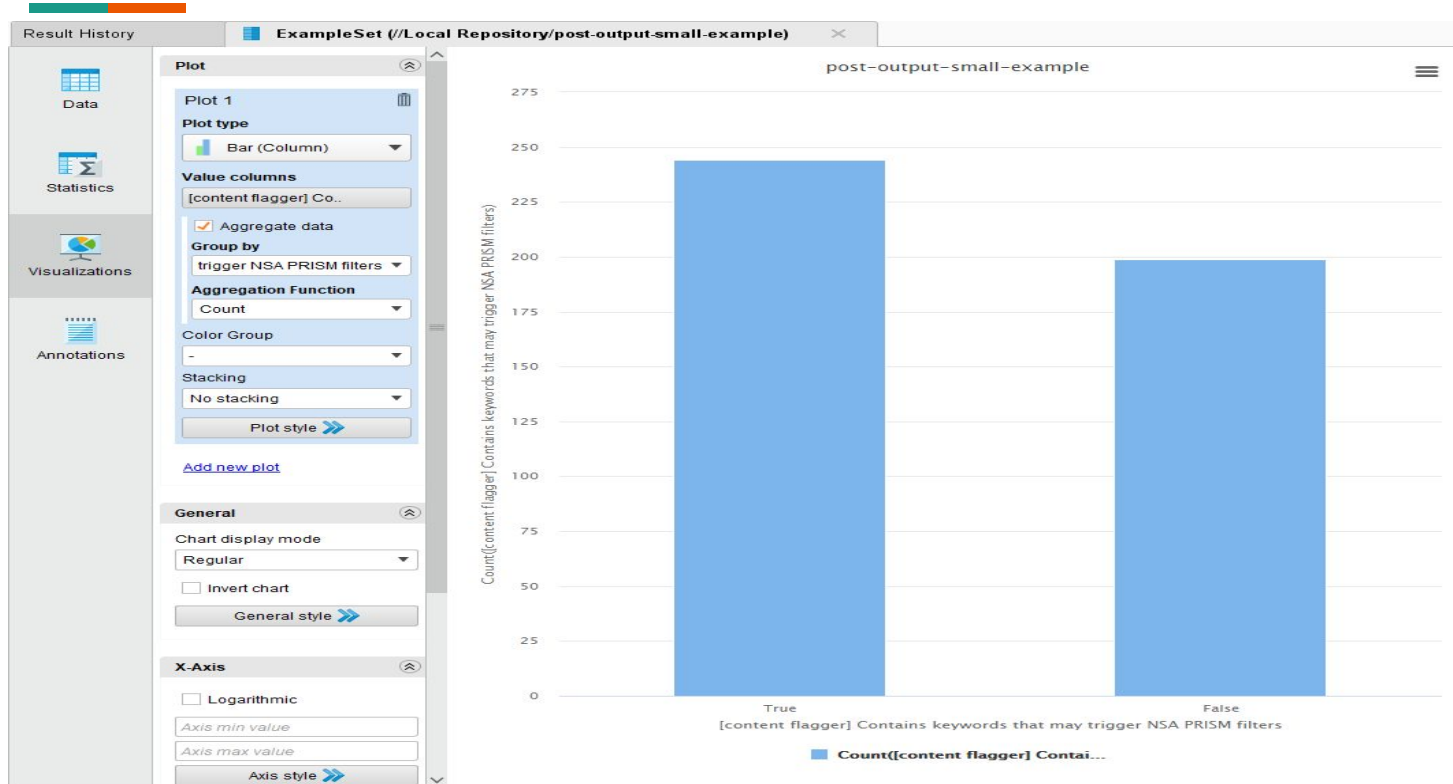
Flaggers /country



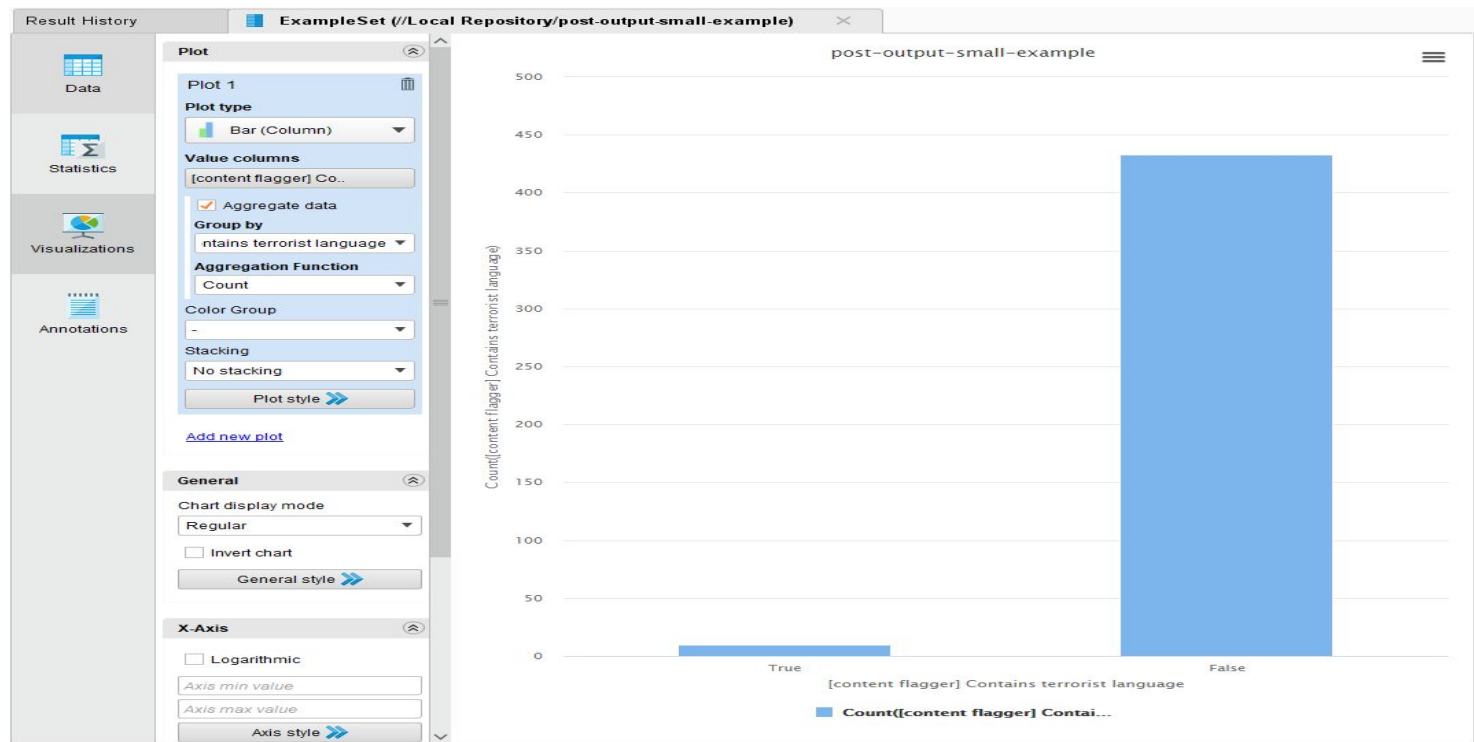
RapidMiner data outcome



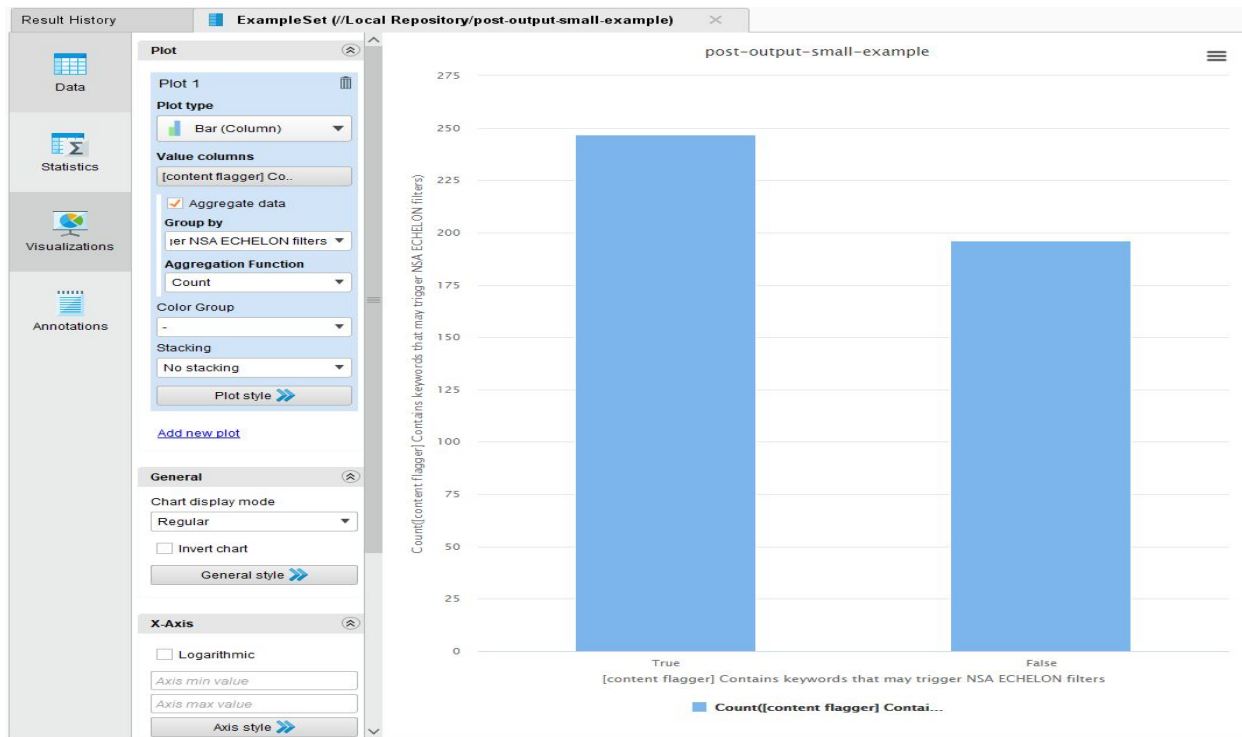
RapidMiner Outcome Cont / content flagger



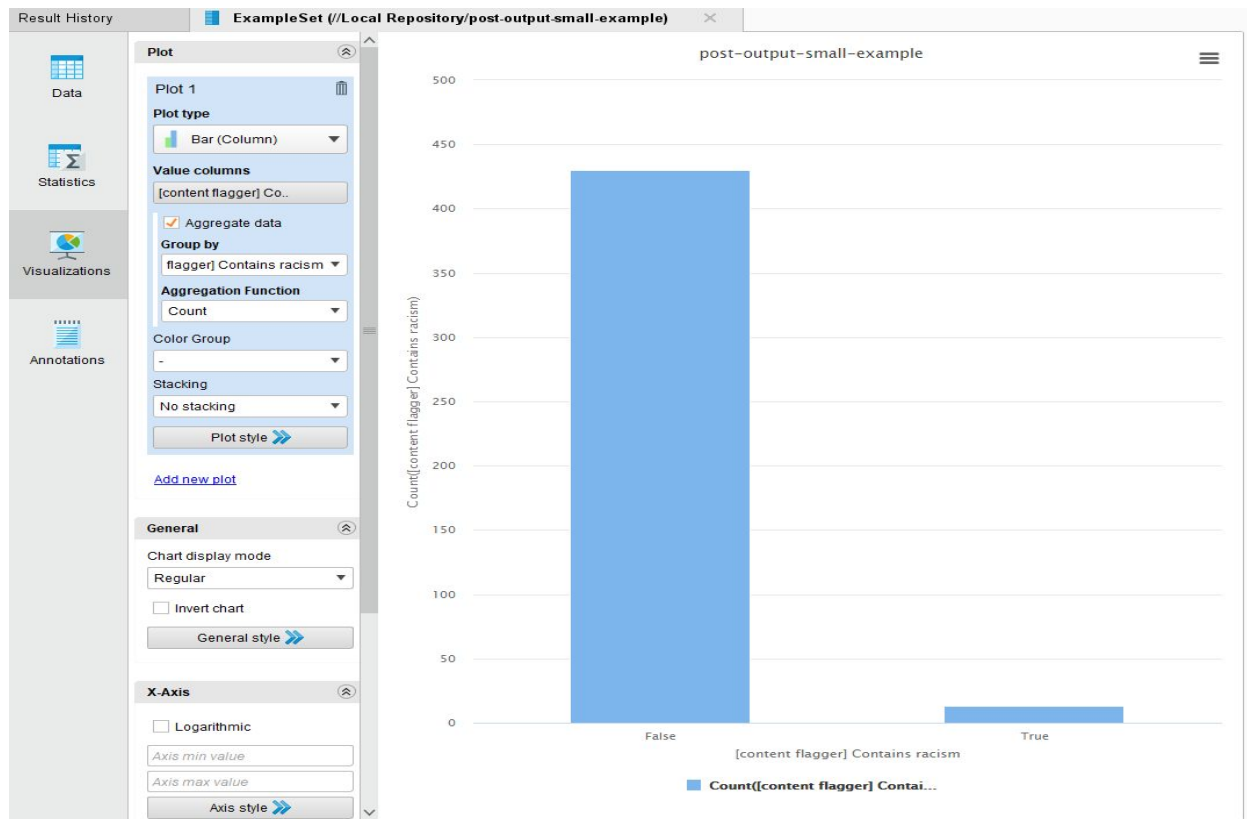
Content / flagger



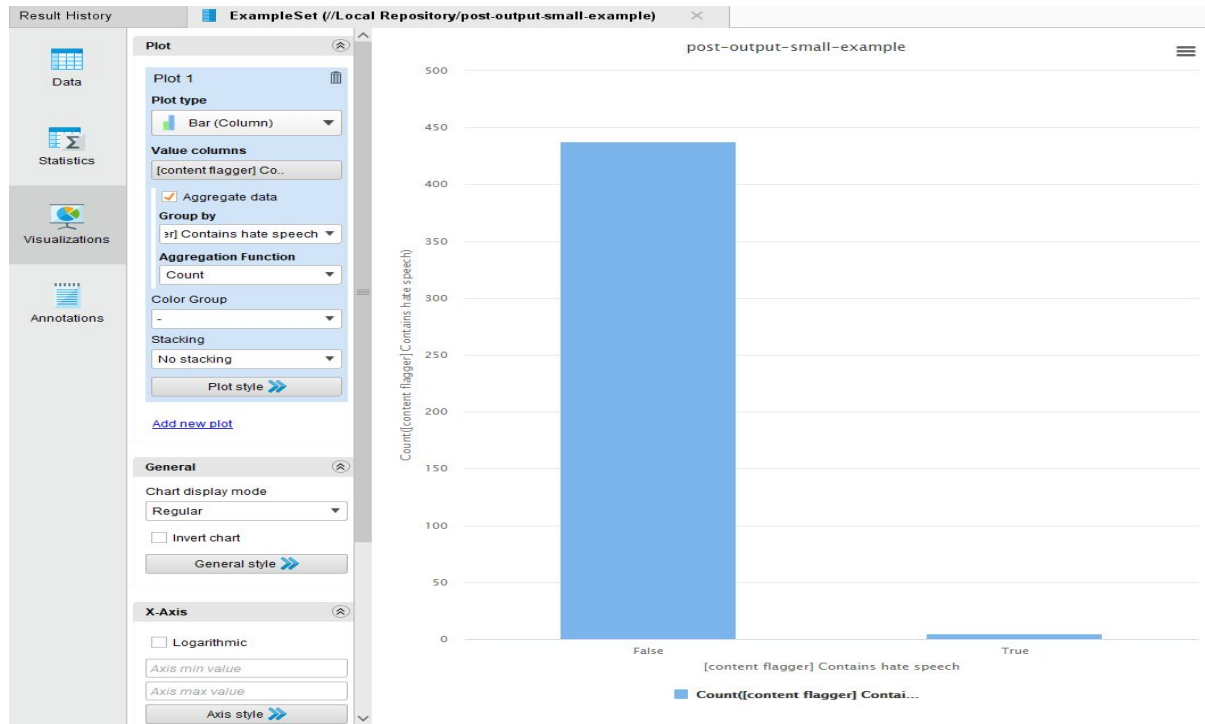
Flagger Cont...



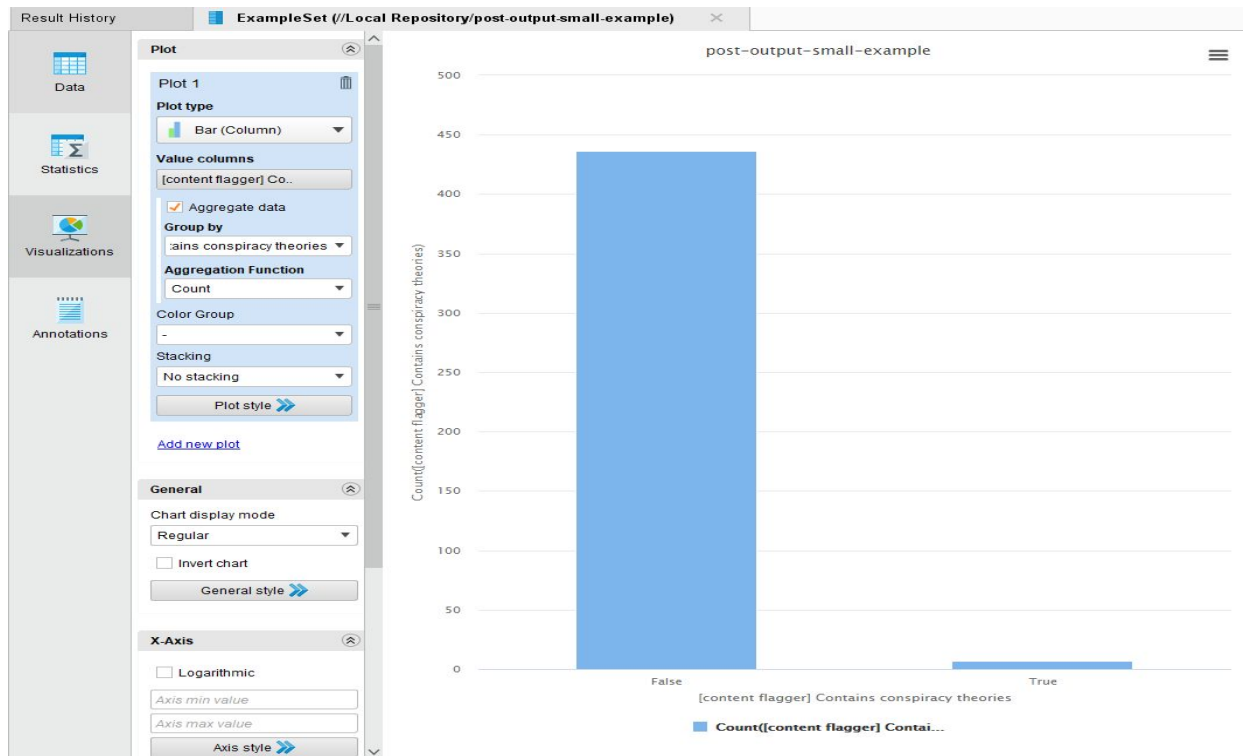
Flagger cont...



Flagger cont...



Flagger / Content





Debugging / Analysis

- Testing of Formatting and Scraping done on Various platforms.
- Testing Scraper Migration from local machines to Public Web
- Used small sample CSV to for migration testing



Debugging / Analysis

Used python Pandas to search through flagged content

- Found 8kun board in DarkWeb

4chan limits metadata collection

- IP Address, Platform, Geo-location, user agent.
- Unsure if VPN is used



Threat Keyword Process

- We started testing with various broad keywords surrounding the topics of terrorism, hate speech, racism, and conspiracies
- After further data analysis, we tweaked our keywords for our terrorism category to resemble The Department of Homeland Security's terrorism keyword searches seen in the below link (Page 23)
 - <https://epic.org/foia/epic-v-dhs-media-monitoring/Analyst-Desktop-Binder-REDACTED.pdf>
- As for our other topics (hate speech, racism, conspiracies) we added on, removed, and renamed keywords to better data after testing iterations

Flagged Content with .onion link

pol/ - Politically Incorrect ▾

4chan 

Index ▾

NSFW

Ghost

Gallery

Stats

Search or insert post number



57KiB, 570x570, seeyouinhell.jpg


[View Same](#)

[Google](#)

[iqdb](#)

[SauceNAO](#)

[Trace](#)

/TIG/ Targeted Individual General Anonymous ID: 1kQj3bwa Thu 28 Nov 22:54:06 2019 No.234619787 

[View](#)

[Reply](#)

[Original](#)

[Report](#)

Quoted By: >>234620256 >>234620985 >>234624101 >>234624251 >>234624433 >>234624490

Old Thread >>> >>232492866

8kun board (still unapproved) <http://jthnx5wyvjvzsxtu.onion/tirevolution/>

New to /TIG/ and gang stalking?

> <https://fightgangstalking.com/tactics-for-fighting-back/>

> <http://www.drrobertduncan.com/dr-robert-duncans-neuropsychological-and-electronic-no-touch-torture-report.html>

> <https://stop007.org/home/faq/>

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

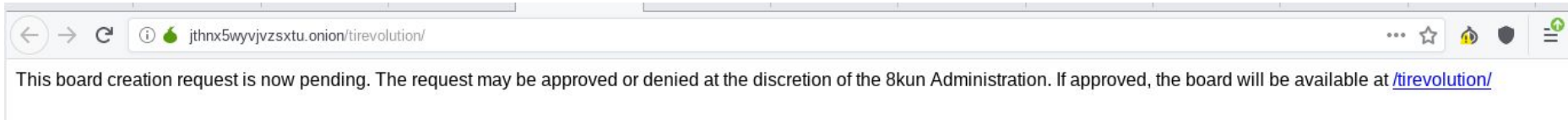
Do your own research. Mind the disinfo. Don't rely on the first google result. Listen to alleged whistle-blowers and victims. There are no survivors. Only victims passed yet. Mind that it's supposed to resemble mental illness. MK Ultra didn't disappear. It evolved.

Old Thread >>> >>232492866

8kun board (still unapproved) <http://jthnx5wyvjvzsxtu.onion/tirevolution/>



Result



Board Request Pending for Approval

As of Dec 2, Link no longer Works



Project Management Structure

Work Breakdown

- Project Management Team
- Developer Team
- Analysis Team



Project Management Structure

Project Stages

- Project Kick-off (Assign Roles and Teams, Set Goals, Scope)
- Code Development
- Training, Testing, Analysis, Further Development
- Project Completion (Final Meetings, Presentation)



Communications

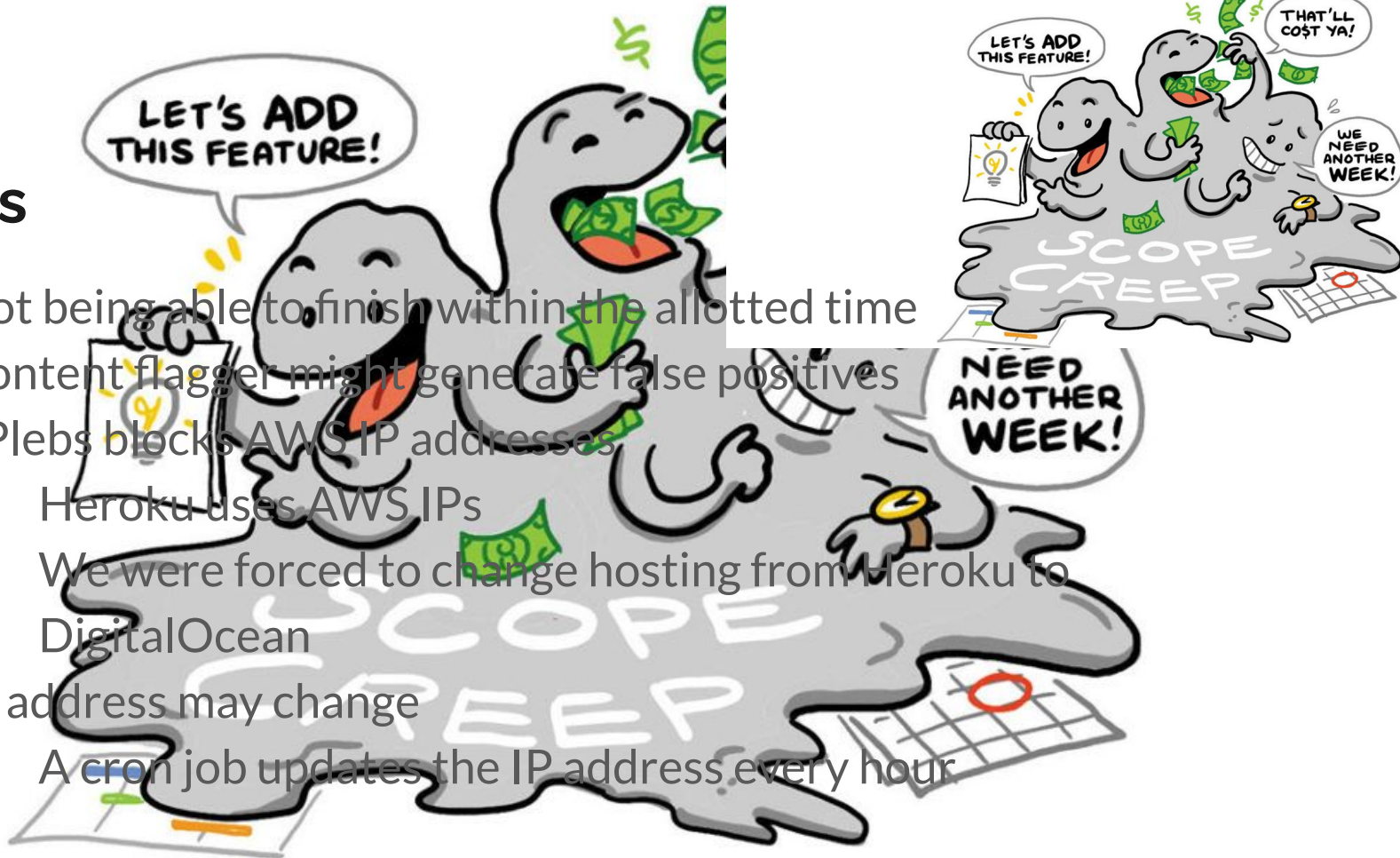
- Main Communication via Google Hangouts
- A Dedicated Google Drive
- GitHub
- Travis

GitHub



Risks

- Not being able to finish within the allotted time
- Content flagger might generate false positives
- 4Plebs blocks AWS IP addresses
 - Heroku uses AWS IPs
 - We were forced to change hosting from Heroku to DigitalOcean
- IP address may change
 - A cron job updates the IP address every hour



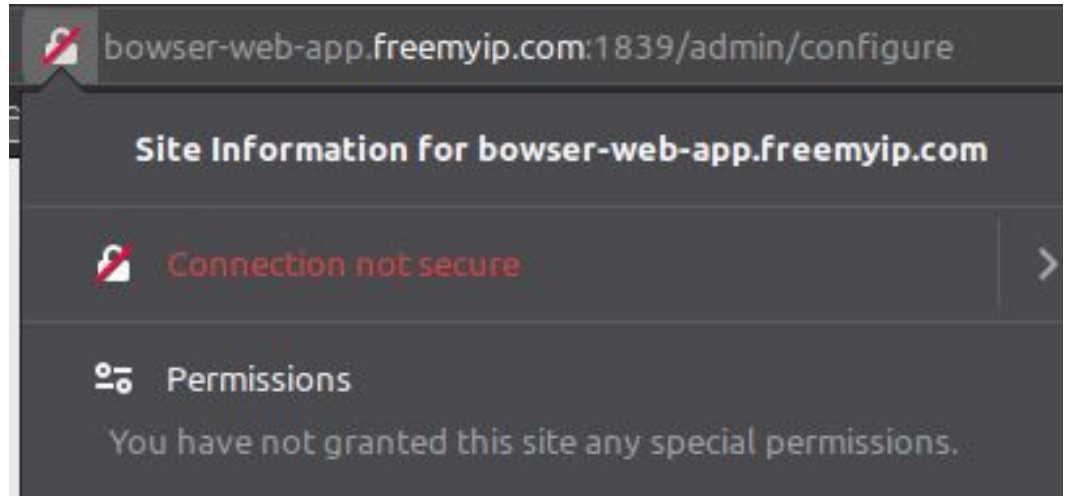


Risks Continued

- Merge conflicts
- TOR sites found
 - Unable to scrape contents of “.onion” sites
- Getting on a Watchlist or Blacklist
 - Flagged Content may contain something illegal

Using HTTP

Currently site does not use
https



Future Work ·





Future Work: Web Interface

- More data sources
 - Can use pushshift.io, a [reddit data gathering](#) tool
 - Simplify quarantined subreddits.
- Creating a more maintainable frontend codebase

pushshift.io

Learn about Big Data and Social Media Ingest and Analysis



Anvil.Works

Full stack python dev



Michael Kotlyar <mkotlyar@hawk.iit.edu>

Nov 19, 2019, 2:32 AM



to sales ▾

Hello,

I am a student attending Illinois Institute of Technology. We are taking a course called ITMS 548 Cybersecurity Technologies. We have a final project where we do open source intelligence gathering and display the data that is user friendly.

So far, we have a project in python that scrapes keywords from message boards that may be considered domestic terrorism.

Is it possible to obtain an educational license where we can host and collaborate with each other for the project?



Meredydd Luff <meredydd@anvil.works>

to me, Anvil ▾

Hi Michael,

Certainly open to the idea - is there something you need that isn't available on the Free plan?

Meredydd

--

Meredydd Luff

Founder - <https://anvil.works>

r/The_Donald



Are you sure you want to view this community?

This community is [quarantined](#)

It is restricted due to significant issues with reporting and addressing violations of Reddit's rules against violence and other aspects of the [Content Policy](#). As a visitor or member, you can help moderators maintain the community by reporting and downvoting rule-breaking content.

Are you certain you want to continue?

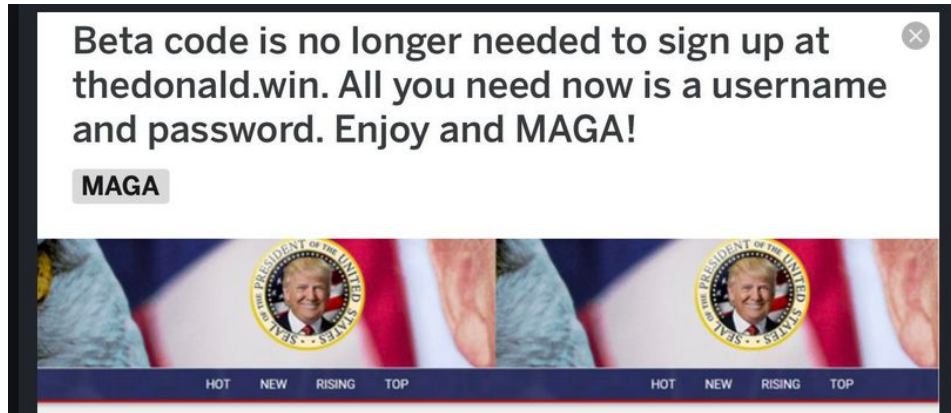
NO THANK YOU

CONTINUE

Future Work: thedonald.win

[Reddit quarantines Trump subreddit r/The Donald for violent comments](#)

[Trump supporters create social media platform tailored for them](#)





Future Work: Web API Backend

- More data sources
- Unified data analysis framework
 - Expose this via the web API
 - Allow users to provide their own custom analysis rules
- Unified 'generic post' framework
 - Combine Facebook, Reddit, Twitter data sources into one format that is universal



Future Work: API Backend

Implement a scalable method of modifying our content flagger database.

- Potentially use existing datasets ('big data') to find more keywords
- Implement Machine Learning / AI to populate and create new filters
-



Future Work: Web API Backend

- There is no authentication on the Web API
 - This is a security problem
 - Our server can be hijacked to send a lot of queries to Reddit or 4Plebs
- Find other Reddit or imageboard APIs to use
 - This cuts down on the code that we must create
 - It also makes us more dependent on other code, which has its own caveats

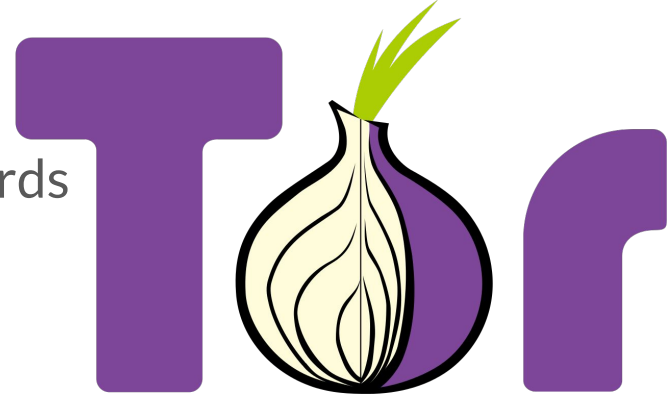
Future Work: The Tor network

Create a wiki of flagged content redirecting to .onion sites

Scrape .onion sites and the Dark Web

Find which Public boards have Dark Web boards

Find the where most-used 8kun is hosted





Future Work: Community and Funding

Leverage Github for feedback, problems, future requests, and contribution. It is Open Source.

Implement a Donation Page for Web Hosting, official Domain, Blockchain Implementation. IPFS/I2P/other peer-to-peer network implementation



InterPlanetary File System (IPFS) - P2P browser



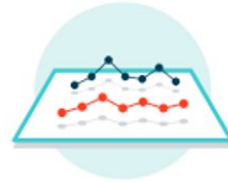
Archivists

It's not enough to organize the world's information—we need to store it in a way the world can remember it. IPFS provides deduplication, high performance, and clustered persistence.



Service providers

If your company delivers large amounts of data to users, a peer-to-peer approach could save you millions in bandwidth. IPFS can provide secure P2P content delivery.



Researchers

If you're working with, distributing, and analyzing huge datasets, IPFS offers fast performance and decentralized archiving.

<https://ipfs.io/>



Future - Angel Investor / Licensing

- Partner with the government to integrate or further develop our tool
- Find private corporate security companies to licence the data to
- Apply an Analyst Model to increase data quality
- Add more data sources to ingest large amounts of data to analyze (such as Twitter)

Closing 1 app and shutting down

To go back and save your work, click Cancel and finish what you need to.



browser.ppt

This app is preventing shutdown.

Thank you! Questions?

Shut down anyway

Cancel