

# Shane's r/memes crawler

Check out the GitHub repository:

<https://github.com/gcshane/reddit-crawler>

Telegram Bot Link:

[https://t.me/shanes\\_reddit\\_crawler\\_bot](https://t.me/shanes_reddit_crawler_bot)



- Front End: React.js + CSS
- Back End: Flask + PRAW + FPDF + Telegram Bot API
- Database: MongoDB

**r/Technology Stack**

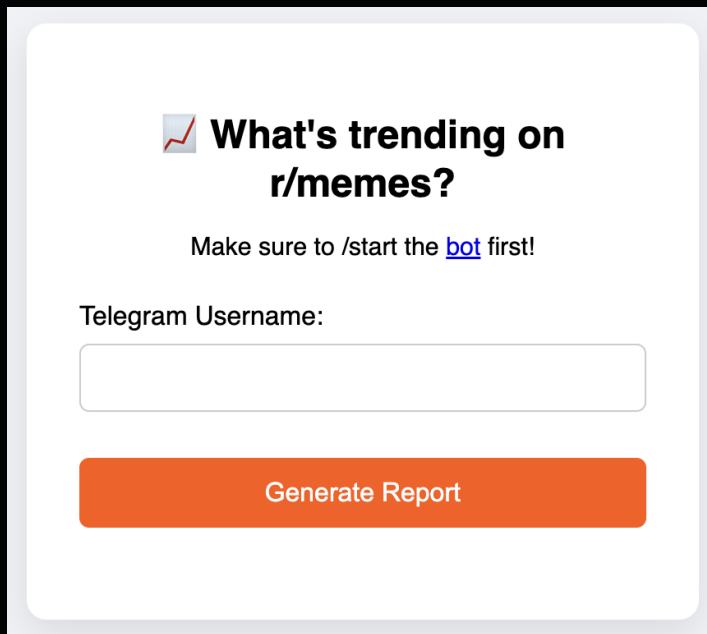
- React.js: wanted to learn something new + widely used
- Flask: familiarity + development speed
- MongoDB: less overhead + no need for multi-table relationship

## **r/Technology Stack Considerations**

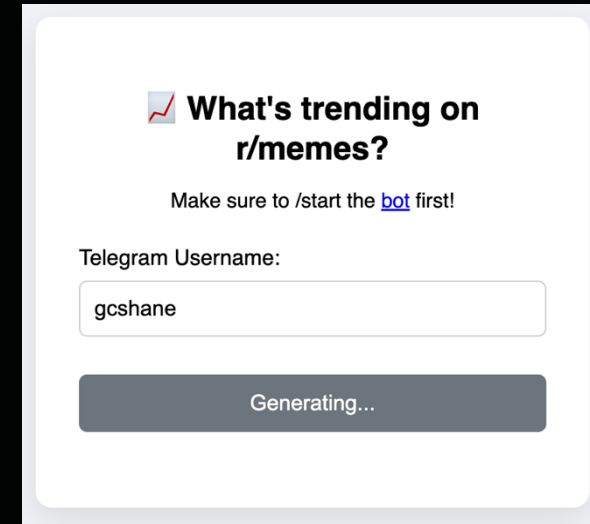
# r/User Flow

2. Redirected to  
Telegram bot and  
/start it

1. Initial View



The initial view is a web form with a white background and a light gray border. At the top, there is a title "What's trending on r/memes?" with a small line graph icon to the left. Below the title, a message says "Make sure to /start the bot first!". Underneath, there is a label "Telegram Username:" followed by an empty text input field. At the bottom, there is a large orange button with the text "Generate Report".




A screenshot of the Telegram bot interface. It shows the title "What's trending on r/memes?" with a line graph icon. Below the title, it says "Make sure to /start the bot first!". There is a label "Telegram Username:" followed by a text input field containing the username "gcshane". At the bottom, there is a gray button with the text "Generating...".

3. Enter Telegram  
Username and click  
"Generate Report"

# r/User Flow

5. User receives report from bot

4. Report generated

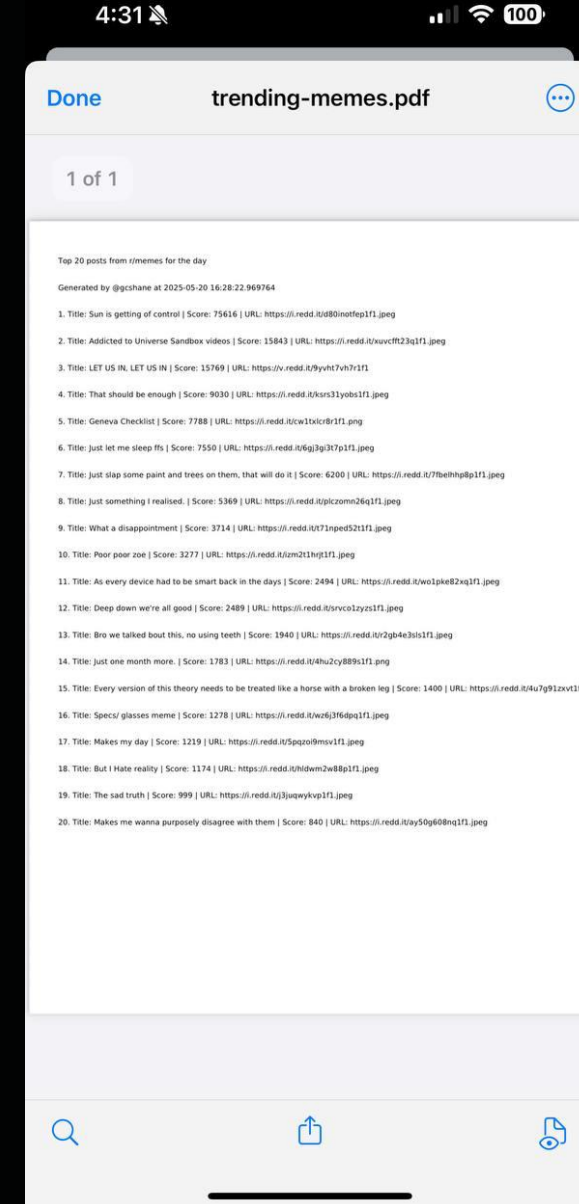
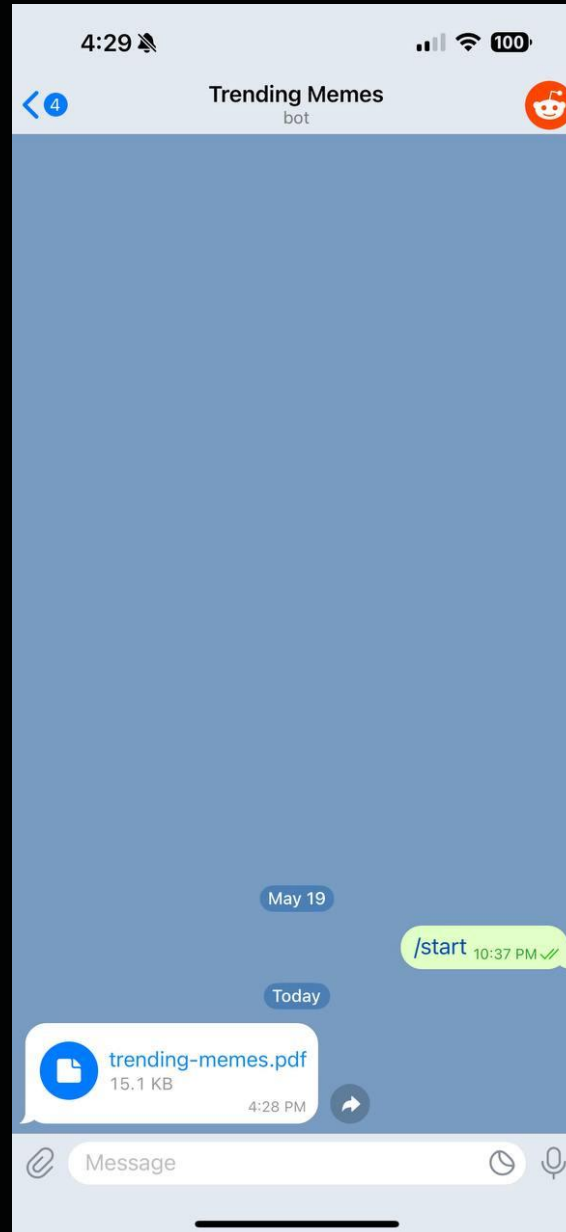
 **What's trending on r/memes?**

Make sure to /start the [bot](#) first!

Telegram Username:

Generate Report

✔ Report successfully sent via Telegram!



6. User can view report

# r/Behind the Scenes

The screenshot displays the Cluster0 MongoDB management interface. At the top, the breadcrumb path is 'SHANE'S ORG - 2023-05-28 > REDDIT CRAWLER > DATABASES'. The Cluster0 logo is on the left, and the version '8.0.9' and region 'AWS Singapore (ap-southeast-1)' are on the right. A navigation bar includes 'Overview', 'Real Time', 'Metrics', 'Collections' (which is active), 'Atlas Search', 'Query Insights', 'Performance Advisor', 'Online Archive', and 'Cmd Line Tools'. Below the navigation bar, it shows 'DATABASES: 1' and 'COLLECTIONS: 1'. On the left sidebar, there is a '+ Create Database' button and a search bar for namespaces. Under the 'reddit\_crawler' namespace, the 'memes' collection is selected. The main panel shows the 'reddit\_crawler.memes' collection with statistics: 'STORAGE SIZE: 20KB', 'LOGICAL DATA SIZE: 3.5KB', 'TOTAL DOCUMENTS: 1', and 'INDEXES TOTAL SIZE: 20KB'. There are tabs for 'Find', 'Indexes', 'Schema Anti-Patterns', 'Aggregation', and 'Search Indexes'. An 'INSERT DOCUMENT' button is in the top right. A query filter bar contains a 'Filter' button, a text input with the query '{ field: 'value' }', and 'Reset', 'Apply', and 'Options' buttons. The query results show 'QUERY RESULTS: 1-1 OF 1' and a single document with the following fields: '\_id' (ObjectId), 'crawl\_time' (ISO 8601 timestamp), and 'items' (Array of 20 items).

Flask application makes API call to Reddit to GET the top 20 posts in the last 24 hours from /memes at the time of crawl and stores the results in MongoDB

# r/AI Assistance

Generally, I had a good idea on what to do from my own research and did most of the coding on my own. AI was mostly used to solve dependency-related issues... and to generate React code which I modified by myself.

Create a simple landing page with a react component that does the following:

- input subreddit name
- input integer for top number of posts
- telegram username
- reddit time filter such as "day"

clicking the submit button should make a http call to my api


Flask app can't detect fpdf2 library

why is my python f string giving me this error UnicodeEncodeError: 'latin-1' codec can't encode character '\u2019' in position 65: ordinal not in range(256)

```
File "/Users/shanegoh/Documents/GitHub/reddit-crawler/venv/lib/python3.13/site-packages/fpdf/fpdf.py", line 469, in add_font
```

```
    raise RuntimeError("TTF Font file not found: %s" % fname)
```

```
RuntimeError: TTF Font file not found: fonts/DejaVuSans.ttf
```



API can be modified to crawl other subreddits

Track recurring trending Reddit users for content creation partnership

Help content creators to understand user engagement trends

## r/What's next?

Alternative Use Cases and Actionalable Insights