Meme Trend Analysis System

Amruta Chaudhari
State University of New York, Binghamton
Binghamton, New York, USA
achaudhari@binghamton.edu

Priyanka Nandwani
State University of New York, Binghamton
Binghamton, New York, USA
pnandwani@binghamton.edu

Narendra Khatpe
State University of New York, Binghamton
Binghamton, New York, USA
nkhatpe@binghamton.edu

Abstract

This proposal outlines the development of a Meme Trend Analysis System designed to collect and analyze meme data from Reddit and 4chan. The system aims to provide insights into meme popularity, template usage, cross-platform dissemination, sentiment context, and cultural impact. By leveraging APIs and custom scrapers, the system ensures continuous data collection, facilitating comprehensive trend analysis essential for understanding digital culture dynamics.

CCS Concepts

• **Social and professional topics**;

Keywords

Data Collection, Meme Analysis, Reddit API, 4chan Scraper, Data Science Pipeline

ACM Reference Format:
Amruta Chaudhari, Priyanka Nandwani, and Narendra Khatpe. 2018. Meme Trend Analysis System. In . ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Memes are a fundamental component of digital culture, serving as vehicles for humor, social commentary, and cultural expression. Analyzing meme trends offers valuable insights into societal shifts, collective sentiments, and the rapid dissemination of ideas across online platforms. This proposal presents the design and implementation of a Meme Trend Analysis System, which will collect and analyze meme data from Reddit and 4chan, laying the foundation for advanced data science projects.

2 Description of Data Sources

2.1 Reddit

- **Description:** Reddit is a community-driven platform where memes frequently emerge and spread across various subreddits.
- **API:** The Reddit API will be accessed using custom-built HTTP requests.
  - `GET /r/subreddit/new.json`: Fetches new posts in a specific subreddit).
  - `GET /r/subreddit/top.json`: Retrieves the top posts in a subreddit over a given time range.
  - `Headers and Authentication`: We'll authenticate using OAuth 2.0
- **Targeted Subreddits:** Subreddits like `/r/memes`, `/r/dankmemes`, `/r/AdviceAnimals`, `/r/PoliticalMeme`, and `/r/PoliticalHumor` will be dynamically added to collect memes.

2.2 4chan

- **Description:** 4chan is an anonymous imageboard that plays a major role in meme creation and dissemination, particularly on boards like `/pol/` and `/b/`.
- **API:** 4chan lacks an official API; thus, a custom-built scraper will be employed to download images and related text from meme-heavy boards.
- **Target Boards:** Boards such as `/b/` (random) and `/pol/` (politically incorrect) will be targeted due to their high meme generation rate.
- **Methods:** GET requests will be utilized to scrape images and associated thread comments from selected boards.

3 Data Collection Plan and System Architecture

The system will continuously collect memes (images and text posts) from Reddit and 4chan. The following components make up the architecture.

3.1 Data Collection Process:

3.1.1 Reddit:

- Custom-built HTTP requests will interact with the Reddit API. We will fetch new and top posts from meme-related subreddits every hour.
- **Authentication:** OAuth 2.0 will be used for Reddit API requests. We will obtain an access token via a POST request to https://www.reddit.com/api/v1/access_token with client credentials.
- We will store the post data (image URLs, metadata, comments) in a database for further processing.

3.1.2 4chan :

- HTTP requests will scrape 4chan by fetching thread information from /b/ and /pol/ boards. We will collect image URLs and text from meme-related threads.
- The scraper will run every 30 minutes to capture new threads and posts.

Meme Trend Analysis System, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

3.2 System Architecture

The architecture can be visualized as follows:

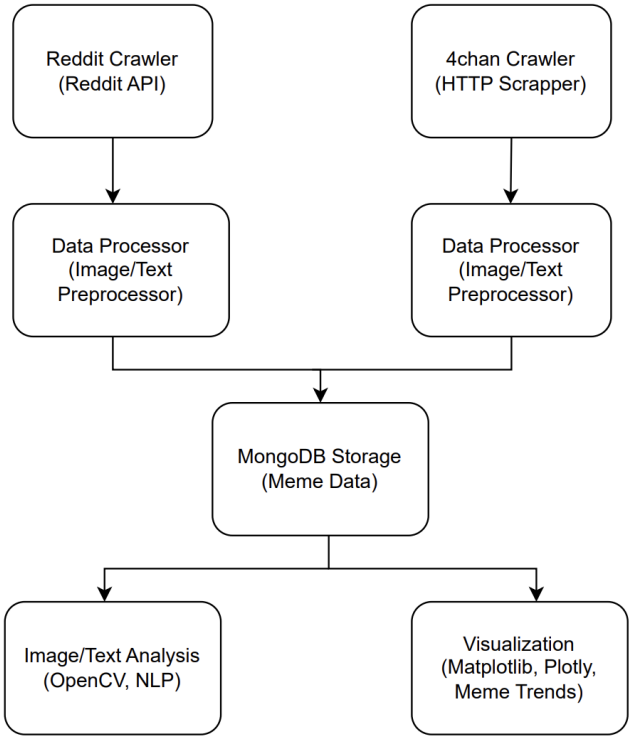

Figure 1: System Architecture Diagram

3.3 Libraries and Tools

- **HTTP Requests:** Custom Python requests to make API calls to Reddit and scrape 4chan.
- **Requests/BeautifulSoup:** For scraping meme images and threads from 4chan.
- **OAuth2.0:** For authenticating Reddit API requests manually.
- **OpenCV:** For image processing and meme recognition (e.g., detecting recurring meme templates).
- **Transformers (Hugging Face):** For analyzing text posts related to memes, identifying context, and sentiment.
- **MongoDB:** To store and manage the collected meme data (images, text, metadata).
- **Pandas/NumPy:** For data processing and analysis.
- **Matplotlib/Plotly:** For visualizing meme trends and metrics.
- **Docker:** For containerization and deployment.
- **Jenkins:** For continuous integration and deployment (CI/CD).
- **Grafana:** For system monitoring and visualization.

4 Measurements and Analysis Ideas

4.1 Meme Popularity Over Time

Track the popularity of specific memes based on post volume, upvote ratios, and comment counts over designated time intervals.

4.2 Meme Template Analysis

Utilize image recognition techniques with OpenCV to identify recurring meme templates. Analyze the most popular templates and their evolution over time.

4.3 Meme Spread Across Platforms

Compare the dissemination rates of specific memes between Reddit and 4chan to understand cross-platform propagation dynamics.

4.4 Context and Sentiment of Meme Posts

Apply Natural Language Processing (NLP) techniques using Transformers to analyze the sentiment and context in which memes are used, categorizing them as positive, neutral, or negative.

4.5 Cultural Impact of Memes

Assess how memes related to current events or societal issues evolve and spread, measuring the speed and reach of memes associated with specific political figures or key news events.

5 Napkin Math

5.1 Reddit

- **Assumptions:**
  - Collecting from 5 meme-related subreddits
  - Using the average number of posts per day for each subreddit for past 30 days.
- **Estimates:**
  - r/PoliticalMeme: 5.30 posts/day
  - r/PoliticalHumor: 32.77 posts/day
  - r/memes: 29.80 posts/day
  - r/dankmemes: 32.63 posts/day
  - r/AdviceAnimals: 20.53 posts/day
  - Total: 121.03 posts/day
  - Total weekly posts = 121.03 posts/day × 7 days/week = 847.21 posts/week

5.2 4chan

- **Assumptions:**
  - Targeting 2 boards: `/b/` (random) and `/pol/` (politically incorrect).
  - Assuming each thread contains an average of 5 image posts.
- **Estimates:**
  - /pol/: 202 threads/day × 5 images/thread = 1,010 images/day
  - /b/: 150 threads/day × 5 images/thread = 750 images/day
  - Total: 1,760 images/day
  - Total weekly images = 1,760 images/day × 7 days/week = 12,320 images/week

5.3 Total Data

- **Reddit:** 847 images/week
- **4chan:** 12,320 images/week
- **Combined:** 13,167 images/week

5.4 Data Size

- Assuming each image (with metadata) is approximately 150 KB.
- Total data size = 13,167 images/week × 150 KB/image = 1,975,050 KB/week ≈ 1.98 GB/week

6 Conclusion

The proposed Meme Trend Analysis System aims to provide a robust framework for collecting and analyzing meme data from Reddit and 4chan. By implementing continuous data collection and leveraging advanced analytical techniques, the system will offer valuable insights into meme popularity, template usage, cross-platform dissemination, sentiment context, and cultural impact. This foundational project is critical for subsequent data science endeavors, ensuring a comprehensive understanding of digital culture dynamics.

References

[1] Reddit API Documentation. Reddit, Inc. Available at: https://www.reddit.com/dev/api/

[2] 4chan API Documentation. 4chan, LLC. Available at: https://github.com/4chan/4chan-API

[3] RFC 6749: The OAuth 2.0 Authorization Framework. IETF. Available at: https://datatracker.ietf.org/doc/html/rfc6749

[4] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools. Available at: https://opencv.org/

[5] Miltner, K. M. (2014). 'There's no place for lulz on LOLCats': The role of genre, gender, and group identity in the interpretation and enjoyment of an internet meme. First Monday, 19(8). Available at: https://firstmonday.org/ojs/index.php/fm/article/view/5391

[6] Shifman, L. (2013). Memes in a Digital World: Reconciling with a Conceptual Troublemaker. Journal of Computer-Mediated Communication, 18(3), 362-377. DOI: https://doi.org/10.1111/jcc4.12013

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30. Available at: https://arxiv.org/abs/1706.03762

[8] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. DOI: https://doi.org/10.18653/v1/2020.emnlp-demos.6