IAmA Data Analysis Live Dashboard Report

Cheng Cui
ccui7@binghamton.edu
Binghamton University
Binghamton, NY

Tianze Liu
tliu76@binghamton.edu
Binghamton University
Binghamton, NY

Mingjie Yan
myan28@binghamton.edu
Binghamton University
Binghamton, NY

1 INTRODUCTION

In this project we build a live dashboard to dynamically display parts of our analysis on r/IAmA data. The dashboard has three tabs, a overview tab for showing the result of all time, a category tab to show the top categories of r/IAmA hosts which corresponds to our first objective in the previous project, and a word cloud tab to show the top frequent keywords in r/IAmA which corresponds to our second objective in the previous implementation. The user can choose a time range by inputting start and end date to query and get the updated charts and figures of that time range accordingly. Thus, through our dashboard users can make sense of the distribution of the information in r/IAmA.

2 INSTRUCTIONS OF DASHBOARD

```
Dashboard/
    Process_Daemon/
        daemon_flask.sh
        killALL.sh
        startDaemon.sh
    scripts/
        Azure.py
        main.py
        Q1.py
        Q2.py
        Watson.py
    static/
        Q1L1.csv
        Q1_new.csv
        Q1_time.txt
        Q2_new.png
        Q2_time.txt
        Q2.png
    templates/
        csv-dataset.html
        form.html
        nav.html
        showpic.html
    tmp/
        fr1.csv
        fr2.csv
        Q1result.txt
        Q2frequency.csv
        Q2result1.txt
    get_input.py
```

We have categorized our files into 5 classes: the scripts folder contains all the codes that handle the back-end process, the templates folder contains all the html files for web page, the static folder contains the updated data and images to be visualized on the web page, the Process_Daemon folder has all the scripts for process daemon and the temp folder stores the intermediate data that are queried from the database. The dashboard can be launched with command: >$ python3 get_input.py under the directory of dashboard.

Below the category and WordCloud tab of the dashboard, users can input the start and end dates in the corresponding text box in the format of "YYYY-MM-DD" and then click the button of "See Time Result" to get the updated chart or figure of the input time range. Before the dashboard update the chart and figure according to user's input, it will take a few seconds to run the back-end data process program to generate the newest results. After the page is fully loaded, the updated chart and figure will be presented below the corresponding tab. If the user input a date that is not fit the format "YYYY-MM-DD", the dashboard will jump to a page with message "Date format is not correct!" as shown in Figure 4

3 DASHBOARD ACCESS REQUIREMENTS

We write a process daemon for our Flask script, it will check the process status every 30 seconds, so that if somehow out Flask framework shutdown, process daemon will restart the services automatically.

3.1 Start Flask microframework

However, if the VM shuts down, process daemon will not work anymore, so here are the steps for manually start the Flask services.

1. Login to our university's internal network by using VPN.

2. Login to VM using SSH with user: myan28 and password: PNcgjEh88.

3. >$ cd /dashboard/Process_Daemon and then >$ bash ./startDaemon.sh .

4. >$ cat looooooog.log to see wheather there is a line shows: "Running on http://0.0.0.0:5000/".

3.2 Access Dashboard

Before you try to access the Dashboard, make sure you already able to access our university's internal network by using VPN. Then try to open http://128.226.28.168:5000/ in any applicable browsers. If you can see the web page as described above, then that's it. If you get something like 'ERR_CONNECTION_TIMED_OUT' meaning failure of access, try steps as follow:

1. Using SSH to open a socket tunnel from your computer to university's internal network. Using command >$ ssh -ND <port> user@ip. This ip can be our CS.remote or can be our host VM

'128.226.28.168' itself. Once you type in the password, do NOT close this ssh window.

2. Change you computer's global proxy setting or browsers' proxy setting, make sure you set sockets to 'localhost:<port>', after that you can using https://www.whatismyip.com/ to check if the tunnel is connected.

3. Once you did connect to university's internal network, you can open http://128.226.28.168:5000/ in any applicable browsers. Now you are supposed to see our Dashboard.

4 RESULT

REFERENCES

[1] https://flask.palletsprojects.com/en/1.1.x/
[2] https://www.chartjs.org/docs/latest/
[3] https://getbootstrap.com/

Live Dashboard for r/IAmA

Overview   Categories   Wordcloud

Overview of Categories and Wordcloud

Last update time: 2020-12-11
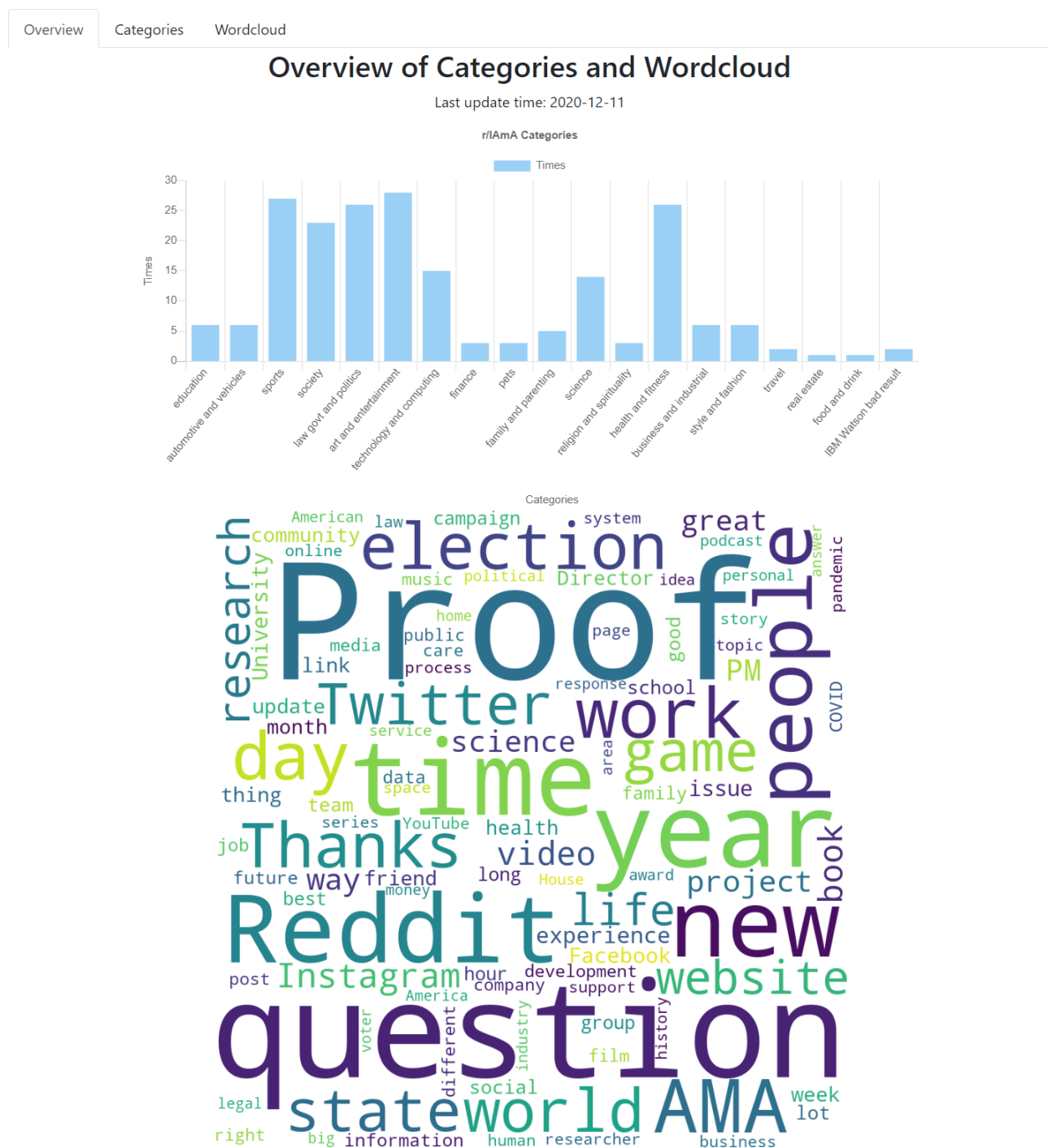
Figure 1: Overview page of DashBoard

Live Dashboard for r/IAmA

Overview | Categories | Wordcloud

Start date YYYY-MM-DD

End date YYYY-MM-DD

See Time Result

Start Time: 2020-11-30 End Time: 2020-12-09
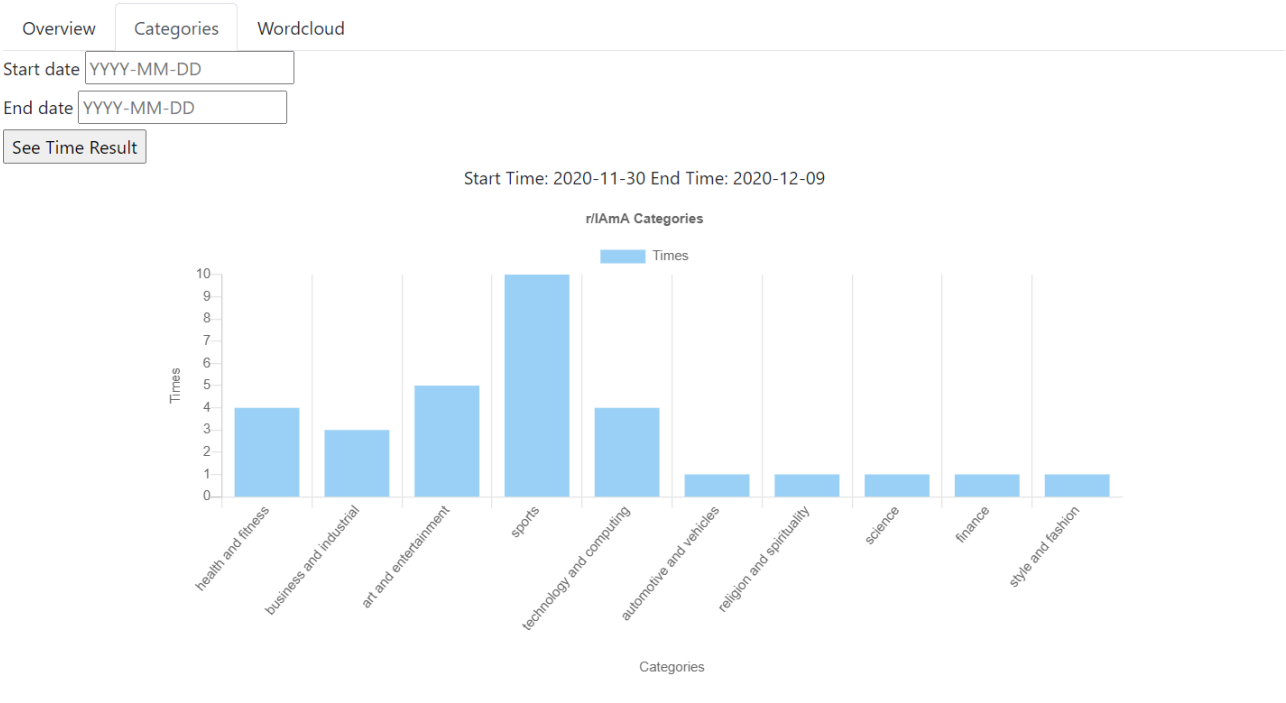
Figure 2: Categories page of DashBoard

Live Dashboard for r/IAmA

Overview   Categories   **Wordcloud**

Start date | YYYY-MM-DD

End date | YYYY-MM-DD

[See Time result]

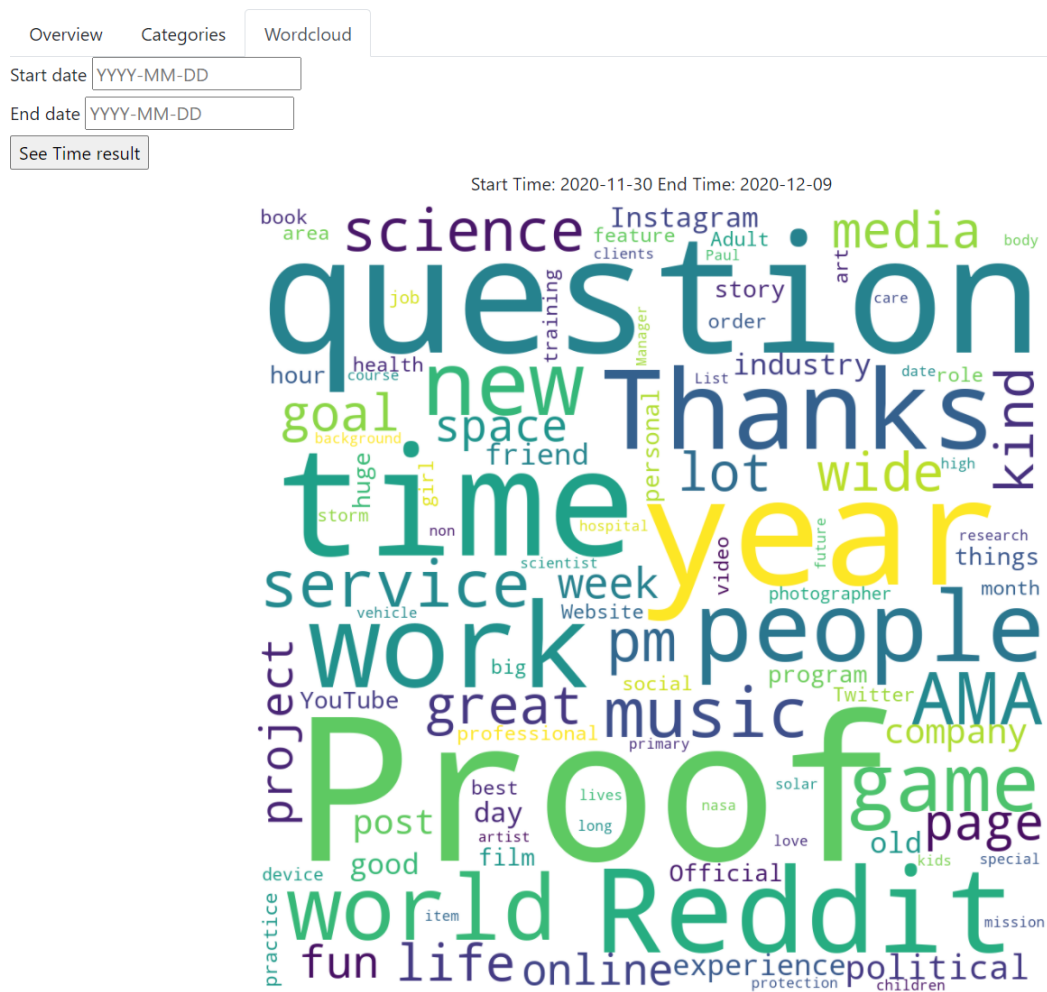Start Time: 2020-11-30 End Time: 2020-12-09

Figure 3: WordCloud page of DashBoard

Date format is not correct!

Figure 4: Error message of incorrect date format