Challenge:

- 1. I went to twitter and got the trends and 10 tweets related to the trends.
- 2. Then I scrapped google search top 10 results with huffington post and the keyword
- 3. I made an android app to display the results.

Programming languages used:

Python-(for twitter trends and scrapping).

PHP-(For calling the python script and getting the results from mysql database).

Android (Java) - an app to show the results of the search.

RDS (MySQL) - to store the results.

Server:

Linux (ubuntu): created a linux server in Amazon Web Services.

Libraries used:

Tweepy (a python library to get the trends and tweets).

Beautiful soup (to parse the text from a web url scrapped).

Tools used:

Putty- to open the terminal for my linux machine in aws.

Winscp-to access the files in my server.

Android studio- to develop the app.

Sublime text- editor that I used to code in python and php.

MYSQL workbench- to access my db.

Approach:

Getting Trends:

- I used tweepy library to get the trends in newyork.
- Once I got the trends I saved in TRENDS table in db

Getting tweets for the trends:

- Once I saved the trends I looped through the trends and searched for tweets related to the trends using tweepy.
- I saved the results in db table TWEETS_TREND with trend_id as a foreign key that references TRENDS table

Getting the HuffPost Urls:

- I scrapped the google search results with query "huffington post"+trend name
- I got the top 10 results and saved them in LINKS_TREND with trend_id as a foreign key that references TRENDS table

Python files:

- Huffpost.py is the main file from which I call other classes to get trends and crawl web
- Trends.py has a class which has different methods to get the trends, tweets, save trends, save tweets, save links etc.
- Google search huffpost.py is the file in which I scrapping the search results of huffington post.

Other Approaches tried:

- I tried to scrap the whole huffington post starting from huffintonpost.com but I encountered a problem i.e I had to wait forever to get the results but other pages that end with .html with in huffington post were easy to scrap but the pages that doesn't end with html I couldn't scrap. I tried to investigate for a while but couldn't find anything. Because of this I had to scrap the google search results.
- I tried using other libraries and different approaches but I settled finally with google search results scrapping.
- You can find all the files in huffpost directory.

PHP Server:

- Once I coded the python files I started coding my php server.
- It contains one file huffpost.php from which I call huffpost.py to refresh the results every time.
- It has two functions which returns json encoded data to my android app once it is opened.
- One function is to return all the trends and tweets.
- Another function is to return all the trends and article links related to trends.

Challenges faced and a note:

- I commented out the auto update of trends because there is limit of calls I can make to twitter api
- Also to mimic the browser to get the google results I am using mechanize library in python and If I frequently get the results google is throwing me a forbidden error

Improvements I can make:

- I can definitely make the scrapping more intelligent and instead of using google search results I could crawl all of huffington post.
- I completed the project in 2 nights. I was also doing my two academic projects so couldn't find much time but completing your challenge was also important for me. So I make compromises in the quality and relevance of results returned.
- I won't abandon the project like this. I will definitely improve the search results and crawling once I am done with my academic projects.
- There can be lot of improvement in the android app too like storing results locally in sqlite database and running the app in background to sync with server data.
- The UI also can be improved using latest material design recycler view and card views instead of list views.

NYU Courses that helped me with the project.

- Mobile app development- To develop android app, Java
- Web search engines- to scrap web and python language I learnt for the course
- Databases course-MySQL, RDS
- Programming languages course- php

Android App:

Activities:

- Helper activity This is the launcher activity. This is where I am calling my php script and getting all the data.
- Main activity-

Once I got all the data I send the user to Main activity.

In Main activity I am using pageviewer with which user can slide right to see all the trends and links related to the trend

The first slide user sees is a list of all trends. If he clicks on any trend it opens into an expandable list and shows related tweets

From the second slide onwards user sees a trend title and list of links related to the trend Once the user clicks on any list. It opens a web browser and he can go through the article If pressed back he will return to the previous link

I have a refresh button on top so he can periodically refresh the results

ScreenShots:





REFRE

HuffPostCodingChallenge

า7s Diner zallgoodproject cjsaw

http://www.huffingtonpost.com/2015/04/16/most-beautiful-homes-2015_n_7079048.html

http://www.huffingtonpost.com/news/thesisproject/

http://www.huffingtonpost.com/nico-pitney/life-lessons-sophia-project_b_6414804.html

http://www.huffingtonpost.com/good-menproject/the-pace-isnt-theproblem_b_6920426.html

http://www.huffingtonpost.com/2015/04/16/ star-wars-jet-r2d2-airplane-787-anaairways_n_7082550.html

http://projects.huffingtonpost.com/dying-tobe-free-heroin-treatment

http://www.huffingtonpost.com/news/the-good-men-project/







REFRE

waySeries Tom7s Diner zallgood

http://www.huffingtonpost.com/2013/10/14/ restaurant-etiquette_n_4086237.html

http://www.huffingtonpost.com/2014/01/13/ jerry-seinfeld-jasonalexander_n_4590374.html

http://www.stereogum.com/1794905/britney spears-will-cover-suzanne-vegas-toms-dineron-giorgio-moroders-deja-vu/news/

http://www.stereogum.com/1703793/hearfall-out-boy-sample-suzanne-vegas-tomsdiner-on-new-single-centuries/mp3s/

http://www.demanjo.com/news/ entertainment/1395360/hear-britney-spearscover-of--tom-s-diner--from-giorgio-moroders-new-album.html

http://www.demanjo.com/news/ entertainment/1395360/first-listen--snippetsof-britney-spears%25E2%2580%2599s-cover-







REFRE

AboutHil... SubwaySeries Tom7s

http://www.huffingtonpost.com/sidney-annestone/http/www.huffingtonpost.com/sidneyanne-stone/second-avenue-subwayseri_b_4080545.html

http://www.huffingtonpost.com/rob-taub/ subway-series_b_1622165.html

http://www.huffingtonpost.com/rob-taub/

http://www.huffingtonpost.com/stevenhirsch/

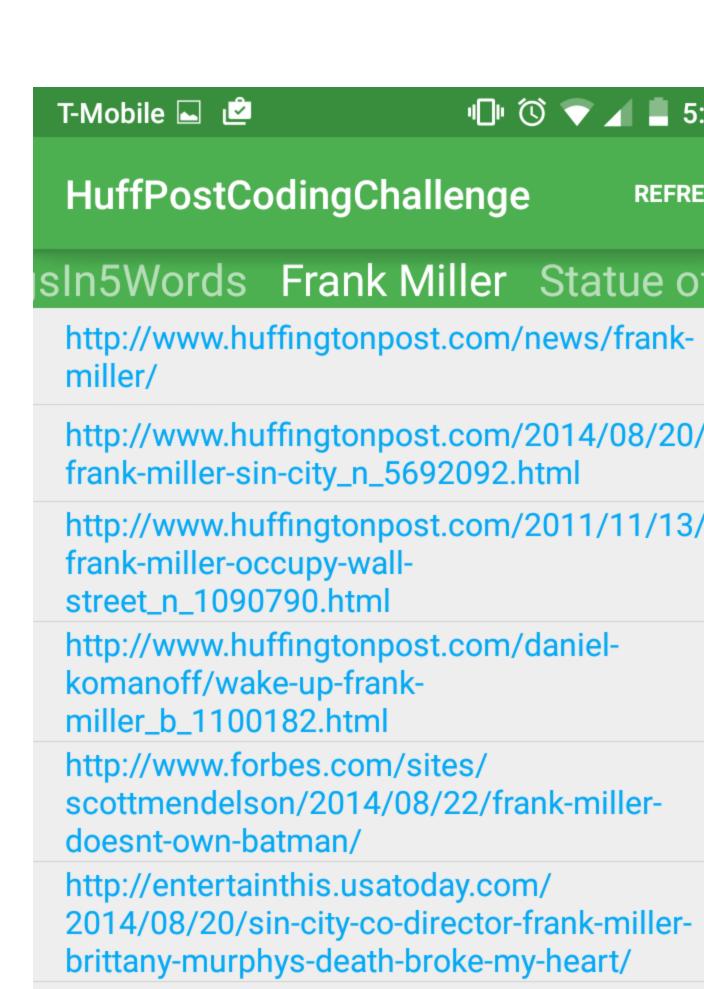
http://videos.huffingtonpost.com/Video/ yankees-to-host-mets-in-game-one-ofsubway-series-518229429

http://www.huffingtonpost.com/news/saltlake-city/

http://www.huffingtonpost.com/david-vines/ http/www.huffingtonpost.com/david-vines/

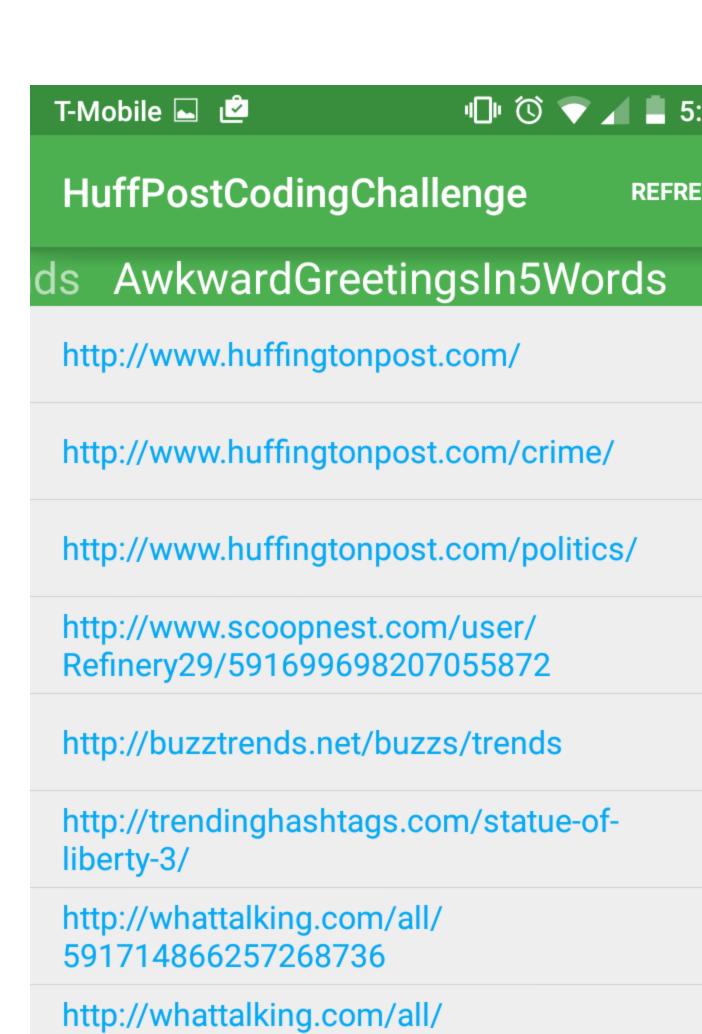


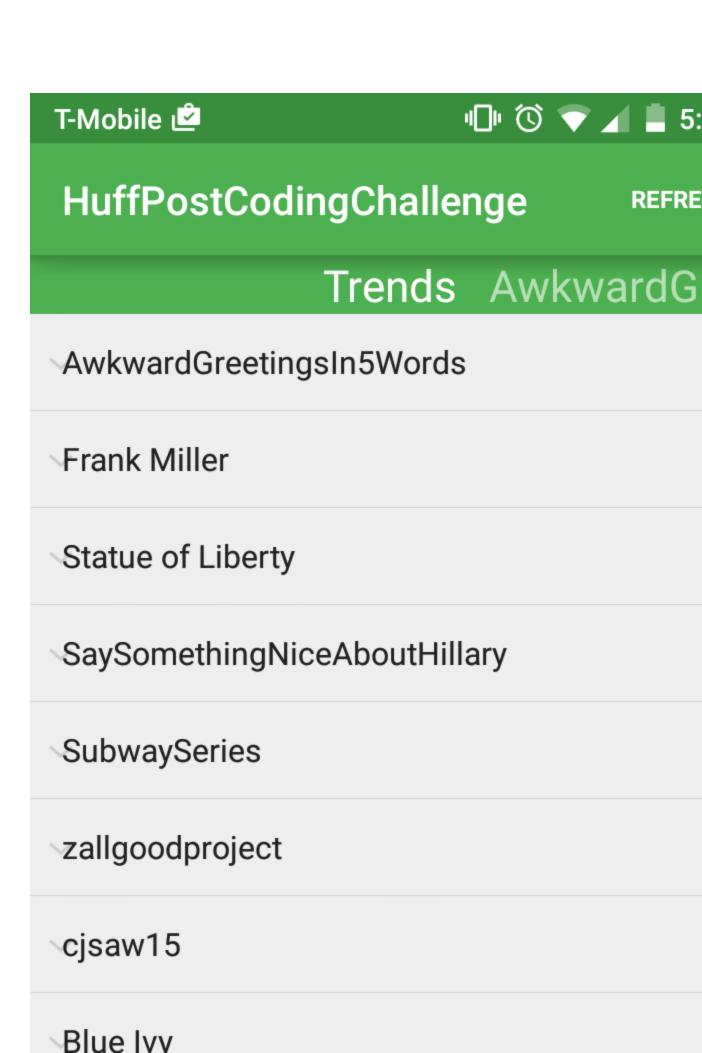
city_n_7041370.html



http://www.thewrap.com/frank-miller-comic-

hook-movies-are-getting-hetter-hut-





REFRE

Trends

SaySomethingNiceAboutHillary

FlashbackFriday

#FlashbackFriday @HayesGrier http://t.co/aT8LOojKod

Another #flashbackfriday of #tourlife 2009 #touring https://t.co/nGTmzuM4dW

RT @ItzGMan713: #FlashBackFriday @blockofrockstudios Workin on Makin Moves Free@JuanGotti @rasheed713 lowg https:// t.co/A0tZCyFwQ1

RT @YRInsider: #FlashbackFriday: #YR @DaytimeEmmys memories with @marchpemberley! http://t.co/eGiitVxTBa

RT @michellemalkin: #FlashbackFriday Neve forget SFC Christopher J. Speer cc



REFRE

Trends

Statue of Liberty

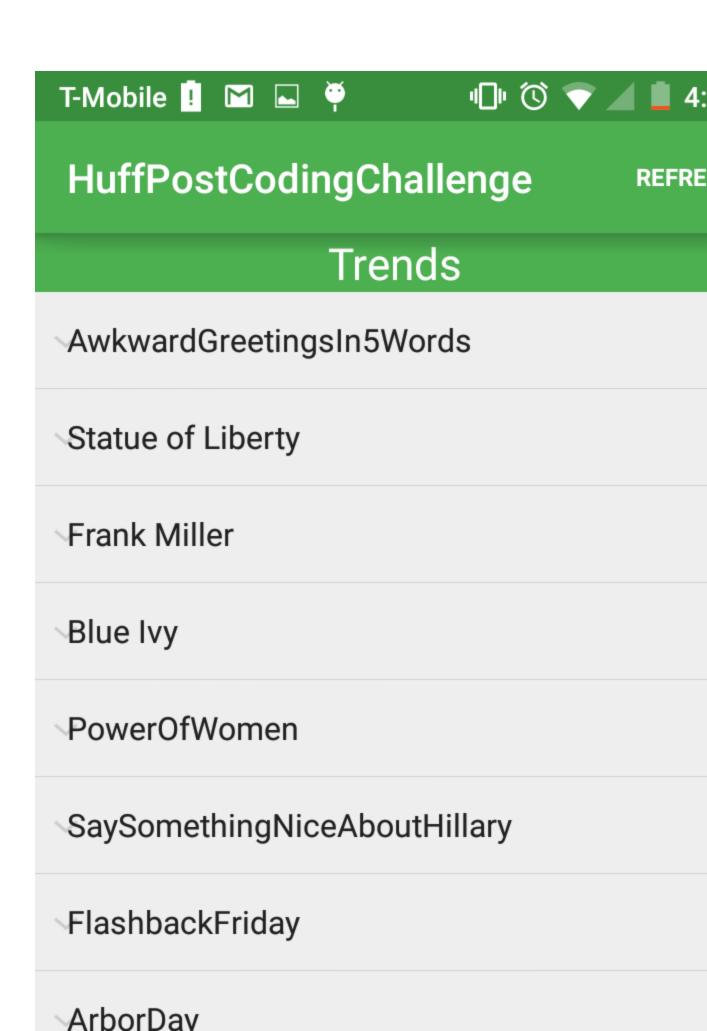
Statue of Liberty deemed safe after bomb threat: media reports: NEW YORK (Reuters) - Law enforcement authoriti... http://t.co/9GfBq18In8

Bomb threat at Statue of Liberty http://t.co/ cxt9930FvX

Statue of Liberty Temporarily Evacuated Over Report of Suspicious Package | TheBlaze http://t.co/soF4hTXEgt

RT @AP: Nothing harmful found after bomb threat led to Statue of Liberty evacuation: http://t.co/wXHssvcclJ

Thorough sweep of Statue of Liberty turns up no explosives http://t.co/N7DzOTCTvm Free tool online http://t.co/J90dT28MyB









THE HUFFINGTON POST

THE INTERNET NEWSPAPER: NEWS BLOGS VIDEO COMMUNITY

