

## **HuffPost Coding Challenge:**

### **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:

# HuffPostCodingChallenge

REFRE

lgoodproject cjsaw15

Blue

<http://www.huffingtonpost.com/>

<http://www.huffingtonpost.com/crime/>

<http://www.huffingtonpost.com/politics/>

<http://www.huffingtonpost.com/world/>

<http://www.huffingtonpost.es/>

<https://twitter.com/buffaloscribe/lists/columbians/members>

<https://twitter.com/atrzop/lists/newsonnews/members>

<https://followerwonk.com/bio/%3Fq%3Djournalism%26l%3D%26p%3D%26frmin>

# HuffPostCodingChallenge

REFRE

n7s Diner   zallgoodproject   cjsaw

[http://www.huffingtonpost.com/2015/04/16/most-beautiful-homes-2015\\_n\\_7079048.html](http://www.huffingtonpost.com/2015/04/16/most-beautiful-homes-2015_n_7079048.html)

<http://www.huffingtonpost.com/news/thesis-project/>

[http://www.huffingtonpost.com/nico-pitney/life-lessons-sophia-project\\_b\\_6414804.html](http://www.huffingtonpost.com/nico-pitney/life-lessons-sophia-project_b_6414804.html)

[http://www.huffingtonpost.com/good-men-project/the-pace-isnt-the-problem\\_b\\_6920426.html](http://www.huffingtonpost.com/good-men-project/the-pace-isnt-the-problem_b_6920426.html)

[http://www.huffingtonpost.com/2015/04/16/star-wars-jet-r2d2-airplane-787-ana-airways\\_n\\_7082550.html](http://www.huffingtonpost.com/2015/04/16/star-wars-jet-r2d2-airplane-787-ana-airways_n_7082550.html)

<http://projects.huffingtonpost.com/dying-to-be-free-heroin-treatment>

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

# HuffPostCodingChallenge

REFRE

waySeries Tom7s Diner zallgood

[http://www.huffingtonpost.com/2013/10/14/restaurant-etiquette\\_n\\_4086237.html](http://www.huffingtonpost.com/2013/10/14/restaurant-etiquette_n_4086237.html)

[http://www.huffingtonpost.com/2014/01/13/jerry-seinfeld-jason-alexander\\_n\\_4590374.html](http://www.huffingtonpost.com/2014/01/13/jerry-seinfeld-jason-alexander_n_4590374.html)

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

<http://www.stereogum.com/1703793/hear-fall-out-boy-sample-suzanne-vegas-toms-diner-on-new-single-centuries/mp3s/>

<http://www.demanjo.com/news/entertainment/1395360/hear-britney-spears-cover-of--tom-s-diner--from-giorgio-moroder-s-new-album.html>

<http://www.demanjo.com/news/entertainment/1395360/first-listen--snippets-of-britney-spears%E2%2580%2599s-cover>

# HuffPostCodingChallenge

REFRE

AboutHil... SubwaySeries Tom7s

[http://www.huffingtonpost.com/sidney-anne-stone/http://www.huffingtonpost.com/sidney-anne-stone/second-avenue-subway-seri\\_b\\_4080545.html](http://www.huffingtonpost.com/sidney-anne-stone/http://www.huffingtonpost.com/sidney-anne-stone/second-avenue-subway-seri_b_4080545.html)

[http://www.huffingtonpost.com/rob-taub/subway-series\\_b\\_1622165.html](http://www.huffingtonpost.com/rob-taub/subway-series_b_1622165.html)

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

<http://www.huffingtonpost.com/steven-hirsch/>

<http://videos.huffingtonpost.com/Video/yankees-to-host-mets-in-game-one-of-subway-series-518229429>

<http://www.huffingtonpost.com/news/salt-lake-city/>

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

# HuffPostCodingChallenge

REFRE

nk Miller Statue of Liberty SaySo

<http://www.huffingtonpost.com/news/statue-of-liberty/>

<http://www.huffingtonpost.com/news/statue-of-liberty-reopening/>

[http://www.huffingtonpost.com/2013/06/27/statue-of-liberty-nyc\\_n\\_3511528.html](http://www.huffingtonpost.com/2013/06/27/statue-of-liberty-nyc_n_3511528.html)

<http://www.huffingtonpost.com/huff-wires/20150424/us--statue-of-liberty-evacuation/>

[http://www.huffingtonpost.com/jonah-green/http://www.huffingtonpost.com/jonah-green/statue-of-liberty-or-lady\\_b\\_849895.html](http://www.huffingtonpost.com/jonah-green/http://www.huffingtonpost.com/jonah-green/statue-of-liberty-or-lady_b_849895.html)

[http://www.huffingtonpost.com/2015/03/05/carowinds-fury-rollercoaster-video\\_n\\_6809700.html](http://www.huffingtonpost.com/2015/03/05/carowinds-fury-rollercoaster-video_n_6809700.html)

[http://www.huffingtonpost.com/2015/04/10/best-walking-tour-new-york-city\\_n\\_7041370.html](http://www.huffingtonpost.com/2015/04/10/best-walking-tour-new-york-city_n_7041370.html)



# HuffPostCodingChallenge

REFRE

sIn5Words Frank Miller Statue o

<http://www.huffingtonpost.com/news/frank-miller/>

[http://www.huffingtonpost.com/2014/08/20/frank-miller-sin-city\\_n\\_5692092.html](http://www.huffingtonpost.com/2014/08/20/frank-miller-sin-city_n_5692092.html)

[http://www.huffingtonpost.com/2011/11/13/frank-miller-occupy-wall-street\\_n\\_1090790.html](http://www.huffingtonpost.com/2011/11/13/frank-miller-occupy-wall-street_n_1090790.html)

[http://www.huffingtonpost.com/daniel-komanoff/wake-up-frank-miller\\_b\\_1100182.html](http://www.huffingtonpost.com/daniel-komanoff/wake-up-frank-miller_b_1100182.html)

<http://www.forbes.com/sites/scottmendelson/2014/08/22/frank-miller-doesnt-own-batman/>

<http://entertainthis.usatoday.com/2014/08/20/sin-city-co-director-frank-miller-brittany-murphys-death-broke-my-heart/>

<http://www.thewrap.com/frank-miller-comic-book-movies-are-getting-better-but->

# HuffPostCodingChallenge

REFRE

## ds AwkwardGreetingsIn5Words

<http://www.huffingtonpost.com/>

<http://www.huffingtonpost.com/crime/>

<http://www.huffingtonpost.com/politics/>

[http://www.scoopnest.com/user/  
Refinery29/591699698207055872](http://www.scoopnest.com/user/Refinery29/591699698207055872)

<http://buzztrends.net/buzzs/trends>

[http://trendinghashtags.com/statue-of-  
liberty-3/](http://trendinghashtags.com/statue-of-liberty-3/)

[http://whattalking.com/all/  
591714866257268736](http://whattalking.com/all/591714866257268736)

[http://whattalking.com/all/  
591714499915599996](http://whattalking.com/all/591714499915599996)

# HuffPostCodingChallenge

REFRE

Trends AwkwardG

✓AwkwardGreetingsIn5Words

✓Frank Miller

✓Statue of Liberty

✓SaySomethingNiceAboutHillary

✓SubwaySeries

✓zallgoodproject

✓cjsaw15

✓Blue Ivy

T-Mobile



4:

# HuffPostCodingChallenge

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 Never  
forget SFC Christopher J. Speer cc

T-Mobile



4:

# HuffPostCodingChallenge

REFRE

## Trends

### ✓Statue of Liberty

Statue of Liberty deemed safe after bomb threat: media reports: NEW YORK (Reuters) - Law enforcement authoriti... <http://t.co/9GfBq18ln8>

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>

T-Mobile



4:

# HuffPostCodingChallenge

REFRE

## Trends

AwkwardGreetingsIn5Words

Statue of Liberty

Frank Miller

Blue Ivy

PowerOfWomen

SaySomethingNiceAboutHillary

FlashbackFriday

ArborDay

T-Mobile



4:

# HuffPostCodingChallenge



# THE HUFFINGTON POST

THE INTERNET NEWSPAPER: NEWS BLOGS VIDEO COMMUNITY

# HuffPostCodingChallenge

REFRE

cjsaw15

Blue Ivy

n

<http://www.huffingtonpost.com/news/blue-ivy-carter/><http://www.huffingtonpost.com/news/beyonce-blue-ivy/><http://www.huffingtonpost.com/news/blue-ivy/><http://www.huffingtonpost.com/news/beyonce-blue-ivy-carter/><http://www.huffingtonpost.com/news/blue-ivy-carter-photos/>[http://www.huffingtonpost.com/2015/01/07/blue-ivy-wardrobe-instagram-photos\\_n\\_6418958.html](http://www.huffingtonpost.com/2015/01/07/blue-ivy-wardrobe-instagram-photos_n_6418958.html)<http://www.huffingtonpost.com/news/blue-ivy-photos/><http://www.huffingtonpost.com/news/blue->



