

Database Team Project

# [ TREND\_CATCHER ]

The Relationship Between Trend and  
K-Drama Frequent Tweets

20102117 Jeonghyun Lee

20102123 Jina Lee

20102125 Yujin Pi

20102128 Yujeong Hwang

# *Contents*



contents



contents



contents



contents



contents



contents



contents



contents



contents

1. Background

2. Framework

3. Description

4. Final result

5. Limitation

## 1. Background

contents

contents

contents

contents

contents

contents

contents

contents

contents

오징어게임 인기에 오징어땅콩 마저...추억의 간식 다시 뜯다

### 오징어게임 인기에 지구촌 '달고나' 열풍...“매대서 품절”

삼양라면, '오징어 게임' 인기에 '펄펄' 끓는다...제2의 '짜파구리'?

오징어게임 뜻밖 호황...오징어 들어간 과자·

라면까지 '불티'

독일도 오겜 열풍...딱지·달고나 제작키트 '불티'



넷플릭스 '오징어 게임' 열풍에 구슬치기 '메카드볼' 뜯다

The 'Squid Game' Costume Effect: White Slip-On Vans Spike 7,800% Since the Series Premiere

'오징어 게임' 덕분...'반스' 흰색 운동화 매출 반짝 증가

'오징어게임 열풍'에 구슬치기 275% 녹색 추리닝 141% 판매 ↑

'오징어게임' 열풍에 코스튬·가면 매출 '껑충'..."핼러윈데이에 딱"

## 1. Background

*Basic idea*



contents



contents



contents



contents



contents



contents



contents

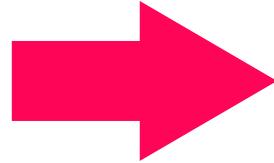


contents



contents

# Drama content



# Industry

*'Is it possible to find the backward linkage effect through Twitter tweets, the world's largest SNS?'*

# 1. Background

## Final target



contents



contents



contents



contents



contents



contents



contents



contents



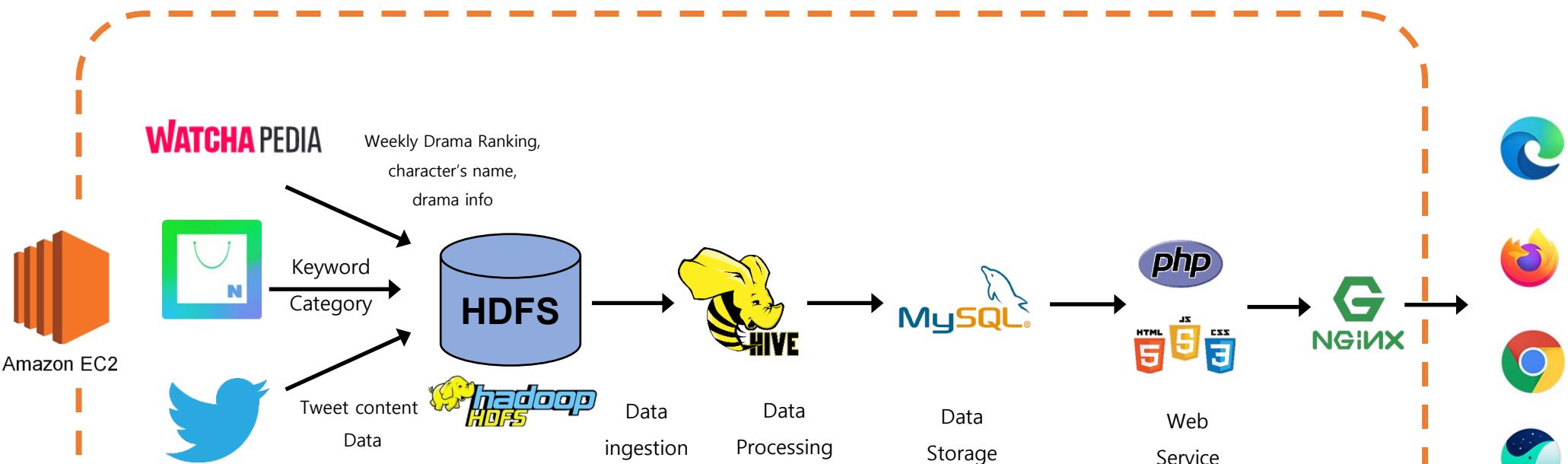
contents



*The final target is to analyze industrial groups and categories that are emerging due to dramas, recommend products that users can purchase through links, and predict the demand of the corresponding industries.*

## 2. Framework

### Our Framework using HDFS



contents

contents

contents

contents

contents

contents

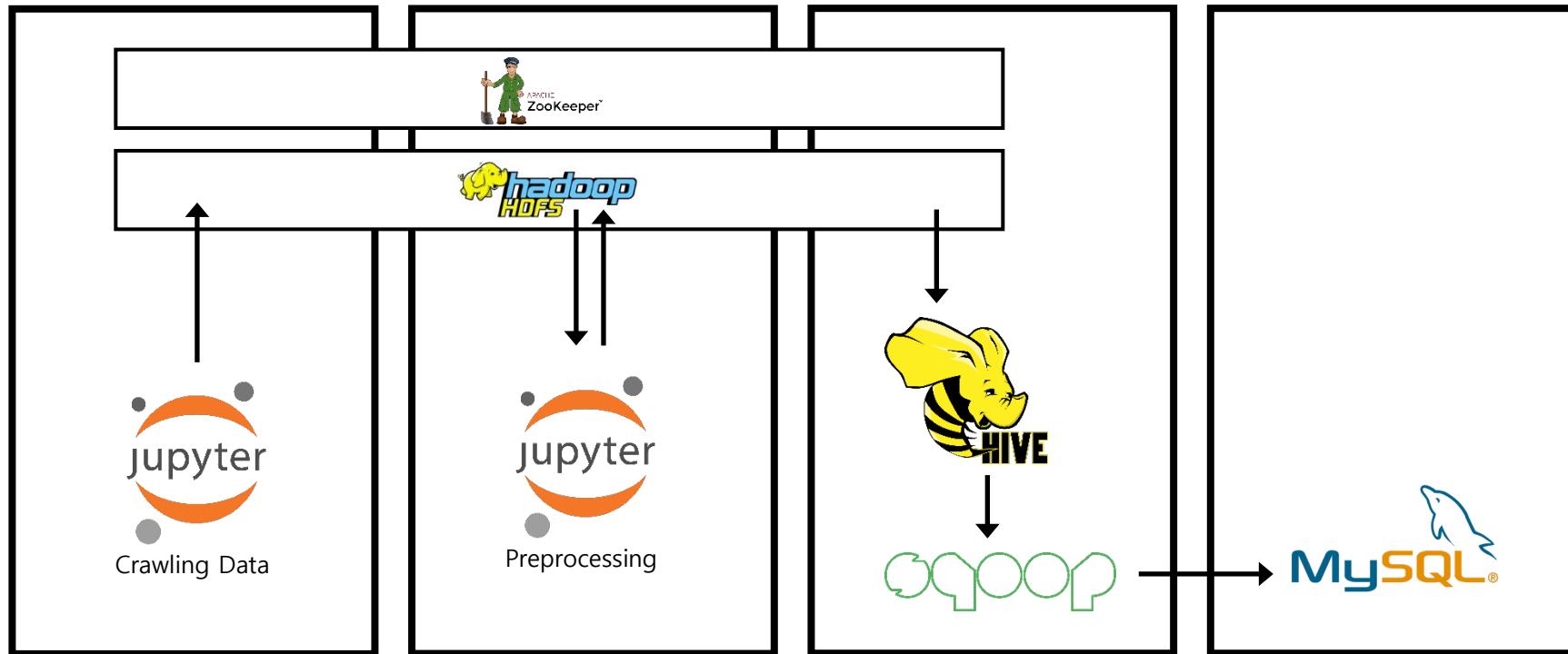
contents

contents

contents

## 2. Framework

### Our Framework using HDFS



- Hadoop: 3.2.2
- Zookeeper: 3.8.0
- Hive: 3.1.2
- Sqoop: 1.4.7
- MySQL: 8.0.29

## 2. Framework

### Our Framework using HDFS

contents

contents

contents

contents

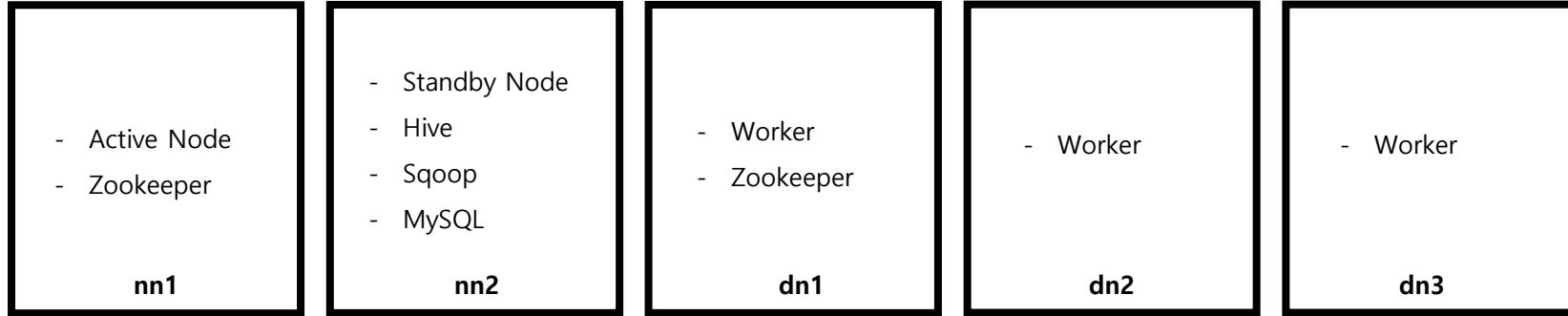
contents

contents

contents

contents

contents



```

ubuntu@nn1:~$ hdfs haadmin -getServiceState namenode1
active
ubuntu@nn1:~$ hdfs haadmin -getServiceState namenode2
standby
ubuntu@nn1:~$

```

```

ubuntu@nn1:~$ jps
11408 JobHistoryServer
11089 ResourceManager
10450 NameNode
11458 Master
10952 DFSZKFailoverController
11577 Jps
10714 JournalNode
ubuntu@nn1:~$

ubuntu@dn1:~$ jps
5505 Jps
5445 Worker
ubuntu@dn1:~$

ubuntu@dn2:~$ jps
4371 worker
4431 Jps
ubuntu@dn2:~$

```

```

ubuntu@nn2:~$ jps
5971 Jps
5752 JournalNode
5930 DFSZKFailoverController
5613 NameNode
ubuntu@nn2:~$

ubuntu@dn1:~$ jps
5299 worker
5359 Jps
ubuntu@dn1:~$

ubuntu@dn3:~$ jps
ubuntu@dn3:~$

```

```

ubuntu@nn1:~$ hdfs fsck /
Connecting to namenode via http://nn1:50070/fsck?ugi=ubuntu&path=%2F
FSCK started by ubuntu (auth:SIMPLE) from /172.31.44.226 for path / at sun May 22 06:40:01 UTC 2022

Status: HEALTHY
Number of data-nodes: 3
Number of racks: 1
Total dirs: 18
Total symlinks: 0

Replicated Blocks:
Total size: 15254313 B
Total files: 31
Total blocks (validated): 30 (avg. block size 508477 B)
Minimally replicated blocks: 30 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)

Erasure Coded Block Groups:
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Ungrouped block groups: 0
Ungrouped corrupt block groups: 0
Ungrouped missing block groups: 0
Ungrouped corrupt missing block groups: 0
Missing internal blocks: 0

FSCK ended at Sun May 22 06:40:01 UTC 2022 in 11 milliseconds

The filesystem under path '' is HEALTHY
ubuntu@nn1:~$

```

## 2. Framework

### Our HDFS & HIVE & MySQL



contents



contents



contents



contents



contents



contents



contents



contents



contents

```
ubuntu@nn1:~$ hdfs dfs -ls /rawdata/rawdata/
Found 13 items
-rw-r--r-- 3 ubuntu supergroup      481 2022-05-22 04:05 /rawdata/rawdata/drama_link.csv
-rw-r--r-- 3 ubuntu supergroup     185 2022-05-22 04:05 /rawdata/rawdata/drama_name.csv
-rw-r--r-- 3 ubuntu supergroup    5208 2022-05-22 04:05 /rawdata/rawdata/dramainfo.csv
-rw-r--r-- 3 ubuntu supergroup 1961562 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-괴이.csv
-rw-r--r-- 3 ubuntu supergroup 623989 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-구될수는없다.csv
-rw-r--r-- 3 ubuntu supergroup 643131 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-그린마더스클럽.csv
-rw-r--r-- 3 ubuntu supergroup 1299864 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-나의해방일지.csv
-rw-r--r-- 3 ubuntu supergroup 401347 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-살인자의쇼핑목록.csv
-rw-r--r-- 3 ubuntu supergroup 2416474 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-스물다섯스물하나.csv
-rw-r--r-- 3 ubuntu supergroup 2346961 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-안나라수마나락.csv
-rw-r--r-- 3 ubuntu supergroup 3006552 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-우리들의블루스.csv
-rw-r--r-- 3 ubuntu supergroup 112800 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-장미연선.csv
-rw-r--r-- 3 ubuntu supergroup 2167097 2022-05-22 04:05 /rawdata/rawdata/ser-tweets-파친코.csv
ubuntu@nn1:~$ |
```

Raw data

```
ubuntu@nn1:~$ hdfs dfs -ls /preprocessing/preprocessing
Found 15 items
-rw-r--r-- 3 ubuntu supergroup      3313 2022-05-22 04:06 /preprocessing/preprocessing/drama1.csv
-rw-r--r-- 3 ubuntu supergroup      3293 2022-05-22 04:06 /preprocessing/preprocessing/drama10.csv
-rw-r--r-- 3 ubuntu supergroup      3314 2022-05-22 04:06 /preprocessing/preprocessing/drama2.csv
-rw-r--r-- 3 ubuntu supergroup      3383 2022-05-22 04:06 /preprocessing/preprocessing/drama3.csv
-rw-r--r-- 3 ubuntu supergroup      3294 2022-05-22 04:06 /preprocessing/preprocessing/drama4.csv
-rw-r--r-- 3 ubuntu supergroup      3284 2022-05-22 04:06 /preprocessing/preprocessing/drama5.csv
-rw-r--r-- 3 ubuntu supergroup      3330 2022-05-22 04:06 /preprocessing/preprocessing/drama6.csv
-rw-r--r-- 3 ubuntu supergroup      3309 2022-05-22 04:06 /preprocessing/preprocessing/drama7.csv
-rw-r--r-- 3 ubuntu supergroup      3363 2022-05-22 04:06 /preprocessing/preprocessing/drama8.csv
-rw-r--r-- 3 ubuntu supergroup      3324 2022-05-22 04:06 /preprocessing/preprocessing/drama9.csv
-rw-r--r-- 3 ubuntu supergroup      481 2022-05-22 04:06 /preprocessing/preprocessing/drama_link.csv
-rw-r--r-- 3 ubuntu supergroup     185 2022-05-22 04:06 /preprocessing/preprocessing/drama_name.csv
-rw-r--r-- 3 ubuntu supergroup     5208 2022-05-22 04:06 /preprocessing/preprocessing/dramainfo.csv
-rw-r--r-- 3 ubuntu supergroup     4573 2022-05-22 04:06 /preprocessing/preprocessing/totalcategory.csv
-rw-r--r-- 3 ubuntu supergroup    39115 2022-05-22 04:06 /preprocessing/preprocessing/totalkeyword.csv
ubuntu@nn1:~$ |
```

Preprocessing data



contents



contents



contents



contents



contents



contents



contents



contents



contents

## 2. Framework

### Our HDFS & HIVE & MySQL

```
hive> show databases;
OK
default
metastore
Time taken: 0.019 seconds, Fetched: 2 row(s)
hive> show tables;
OK
drama1
drama10
drama2
drama3
drama4
drama5
drama6
drama7
drama8
drama9
drama_link
drama_name
dramainfo
totalcategory
totalkeyword
```

Hive data

## 2. Framework

### Our HDFS & HIVE & MySQL

contents

contents

contents

contents

contents

contents

contents

contents

contents

The screenshot displays three separate MySQL query results side-by-side, each showing data from a metastore table:

- Left Query:** `SELECT * FROM metastore.drama1;` This table contains a list of dramas with columns: keyword, count, link, and category. Examples include 김자기 (33), 경기도 (59), 구자경 (46), 넷플릭스 (47), 드라마 (952), 등장인물 (32), 라이프 (31), 마지막 (35), 메이킹 (100), 벨풀별 (63), 블루스 (284), 비하인드 (32), 스타일 (28), 스틸컷 (75), 서정률 (52), 시청자 (27), 아버지 (48), 아저씨 (224), 어게인 (47), 이야기 (94), 재방송 (29), 주인공 (32), 출판/음악 (33), 커넥터 (86), 캐스팅 (25), 파진코 (29), and 포스터 (102).
- Middle Query:** `SELECT * FROM metastore.dramainfo;` This table contains a list of drama titles with columns: name, info, and poster. Examples include 구불수는없다, 그린마더스클럽, 나의해방법지, 살인자의소평목록, 스물다섯스물하나, 안나라수마니라, 우리들의블루스, 장미연선, and 파진코.
- Right Query:** `SELECT * FROM metastore.totalcategory;` This table contains a list of categories with columns: rank, category, and keyword. Examples include 상황/건강, 패션의류, 드라마, 가구/인테리어, 식품, 출산/육아, 도서, 인테리어조명, 육설정리랑/정리함, 대파, 디자털/가전, 화장품/미용, 스포츠/레저, 골드바, 피규어, 로풀, 패션반지, 플라스틱용기, 가정용거일기, 장식인형, 옷장, and 젤리.

### 3. Description

#### Watchapedia & Naver Crawling

contents

contents

contents

contents

contents

contents

contents

contents

contents

```
def crawling(maxTweets):
    headers = {
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/73.0.3683.86 Safari/537.36'}
    data = requests.get('https://pedia.watcha.com/ko-KR/?domain=tv', headers=headers) # Watchapedia
    soup = BeautifulSoup(data.text, 'html.parser')

    dramas = soup.find('ul', attrs={'class': 'css-yqjebm-VisualUL'}).find_all('li')
```

Crawling Korean TOP10 Drama Ranking and Character Name Required for  
Preprocessing

```
url = "https://search.shopping.naver.com/search/all?query=" + str(input) + "&cat_id=&frm=NVSHTAC"

data = requests.get(url, headers=headers)
soup = BeautifulSoup(data.text, 'html.parser')

#     print(url)
TVRank = soup.find('span', attrs={
    'class': 'filter_text_over_3zD9c'}).get_text()
```

Crawling categories that correspond to keywords in Naver Shopping

### 3. Description

#### Twitter Crawling

♥ contents

💬 contents

✈ contents

🔖 contents

📝 contents

❗ contents

📎 contents

📝 contents

📝 contents

```
maxTweets = 180

# Creating list to append tweet data to
for i in range(len(drama_name_list)):
    # init list and dataframe
    tweets_list1 = []
    tweets_df1 = pd.DataFrame()

    scrape = sntwitter.TwitterSearchScraper(drama_name_list[i] + ' since:2022-05-01 until:2022-05-21').get_items()

    # Using TwitterSearchScraper to scrape data
    for j, tweet in enumerate(scrape):
        if j >= maxTweets:
            break
        tweets_list1.append([tweet.date, tweet.id, tweet.content, tweet.user.username])

    # Creating a dataframe from the tweets list above
    tweets_df1 = pd.DataFrame(tweets_list1, columns=['Datetime', 'Tweet Id', 'Text', 'Username'])
```

Crawling keywords so far for the corresponding drama on Twitter

### 3. Description

#### Morpheme Preprocessing

contents

contents

contents

contents

contents

contents

contents

contents

contents

```

for i in range(len(drama_name_list)):
    name = drama_name_list[i]
    df = pd.read_csv('drama_csv/user-tweets-{}.csv'.format(name))
    df_text = df.iloc[:, [2]] # type : DataFrame (Only text columns are loaded)
    list_text = np.array(df_text.values).flatten().tolist() # type : list (convert 'df' to 'list')
    for k in range(len(list_text)):
        try:
            result1 = re.match("^.*(?=\\()", list_text[k]).group()
        except AttributeError:
            result1 = re.match("^.*(?=\\()", list_text[k])
        # result1 = translator.translate(result1, dest='ko')
        if str(result1) != 'None':
            list_text[k] = str(result1)

        result = ' '.join(s for s in
                          list_text) # type : String (Transform 'list_text' into a string type
                               # 'result' for morpheme analysis into a single string)
        tokens = Okt().nouns(result) # morpheme analysis list
    
```

After processing the special characters, only the nouns are extracted

```

text = nltk.Text(new_tokens)
topWord = text.vocab().most_common(30) # [(top1keyword, count),(top2keyword, count),...]
    
```

Keywords are selected up to 30th

### 3. Description

#### Drama Character Preprocessing



contents



contents



contents



contents



contents



contents



contents



contents



contents

```
for c in characters:
    character = c.find('a').attrs.get('title')

    # Name of person in drama / actual name is separated and saved as a list
    character_list = []
    character_convert = character.replace(" ", "").replace("( ", "").replace(" )", "").replace("출연", " ").replace(
        "|",
        "").replace(
        "감독", "").replace("극본", "").replace("원작", "")
    character_list = character_convert.split()

    if len(character_list) == 2:
        real_name_list.append(character_list[0])
        character_name_list.append(character_list[1])
    elif len(character_list) == 1:
        real_name_list.append(character_list[0])
    else:
        continue
```

The character's name and real name are drawn from the drama

### 3. Description

#### Web programming UX/UI

contents

contents

contents

contents

contents

contents

contents

contents

contents



In main page, it contains top 10 movies ranking, keywords from the entire drama to show the overall keyword ranking and the best category (item). It also contains three most frequent categories and five items about them

### 3. Description

#### Web programming UX/UI

♥ contents

💬 contents

↗ contents

↙ contents

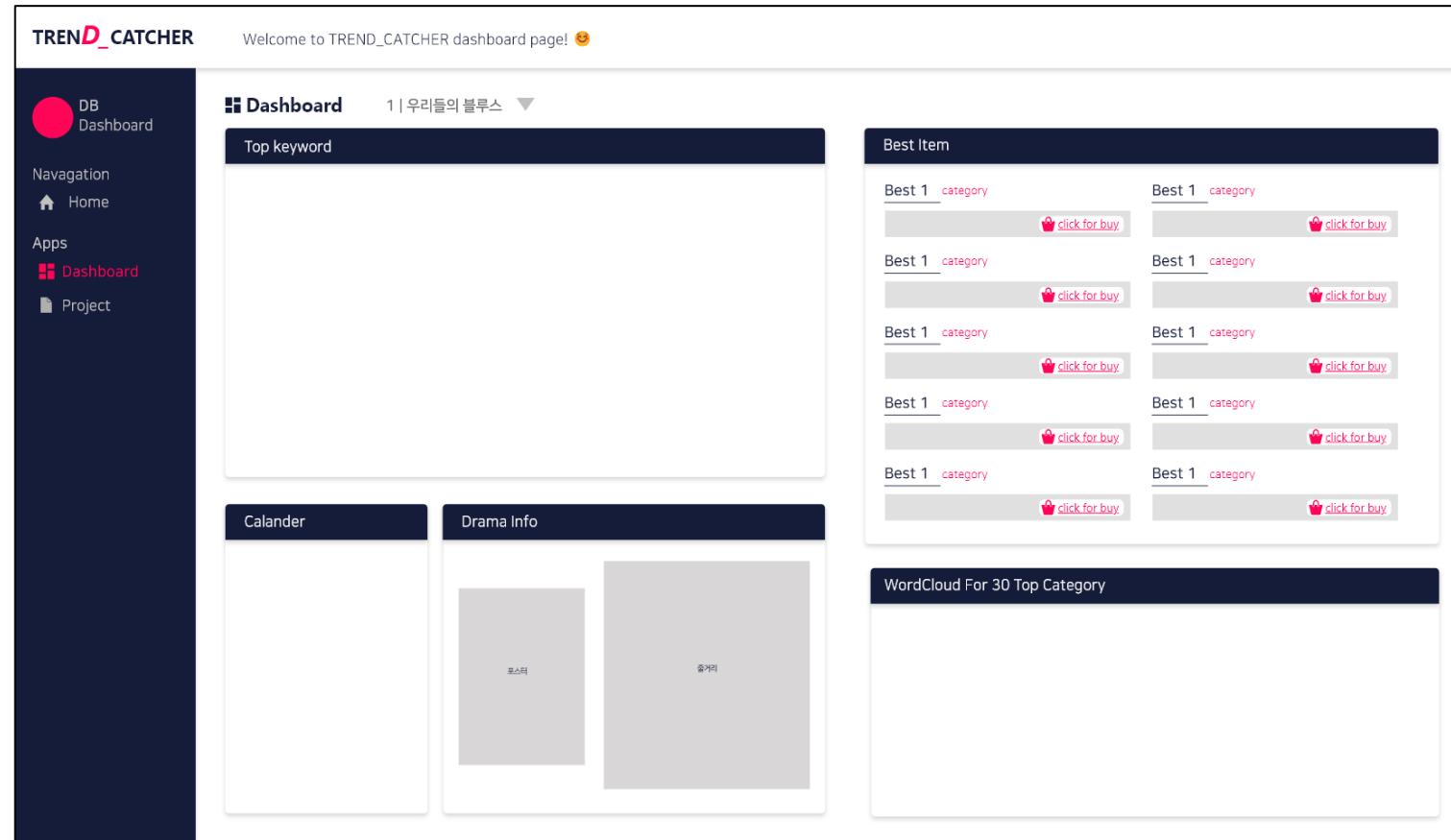
(contents)

contents

contents

contents

contents



In dashboard page, it contains top keyword ranking about each dramas, best 10 categories, drama information and wordcloud for 30 top categories.



contents



contents



contents



contents



contents



contents



contents



contents



contents

### 3. Description

*Web programming – front using html & css*

```

File Edit Selection View Go Run Terminal Help
EXPLORER main.html Dashboard.html Project.html main.js
GROUP 1
main.html
Dashboard.html
Project.html
main.js
GROUP 2
main.css
dashboard.css
project.css
box.css
fulcalendar-sche...
examples
lib
locales
locales-all.js
locales-all.min.js
main.css
main.js
main.min.css
main.min.js
LICENSE.txt
README.md
img
211676_home_ic...
211676_home_ic...
페스 4.png
페스 4.svg
페스 4.png
basket-buy-cart...
g1580.png
g1580.svg
logo.pptx
js
main.js
Dashboard.html
Project.html
OUTLINE
TIMELINE
main.html - 프로젝트 - Visual Studio Code
main.html
main.css
project.css
box.css
navbar&footer.css
project.css
project_text1 {
position: absolute;
top: 262px;
left: 436px;
width: 409px;
height: 100px;
}
project_text1 {
text-align: left;
font: normal normal bold 30px/34px NanumSquare_ac;
letter-spacing: 0px;
color: #151C3A;
opacity: 1;
}
.project_text2 {
text-align: left;
font: normal normal bold 20px/22px NanumSquare_ac;
letter-spacing: 0px;
color: #151C3A;
opacity: 1;
}
section1 project_right {
margin-left: 28px;
width: 826px;
height: auto;
border-radius: 5px 5px 0px 0px;
opacity: 1;
background-color: transparent;
}
.neivy_box {
top: 163px;
left: 892px;
width: 826px;
height: 437px;
background: #151C3A 0% 0% no-repeat padding-box;
box-shadow: inset 3px 0px 6px #00000029, 0px 3px 6px #00000029,
border-radius: 10px;
}

```

By using various html tags and Java script we have learned in the database lecture,  
we made three separate pages



contents



contents



contents



contents



contents



contents



contents



contents



contents

### 3. Description

Web programming – back using php & javascript

The screenshot shows a code editor window titled "main.php — php". The left sidebar displays a file tree for a "PHP" folder containing files like .vscode, css, fullcalendar-scheduler, img, js, server, conf.php, drama1.php, drama2.php, drama3.php, drama4.php, drama5.php, drama6.php, drama7.php, drama8.php, drama9.php, drama10.php, and favicon.ico. The main editor area contains the following PHP code:

```
<?php
$conn = mysqli_connect(
    '3.35.218.159',
    'hive',
    'hive',
    'metastore');

$sql = "SELECT * FROM dramainfo";

$result = mysqli_query($conn, $sql);

$names = array();
while($row = mysqli_fetch_array($result)) {
    array_push($names, $row['name']);
}

///////////////////////////////
$sql_t = "SELECT * FROM totalkeyword";
$result_t = mysqli_query($conn, $sql_t);

$items = array();
$categories = array();
$links = array();
$counts = array();

while($row_t = mysqli_fetch_array($result_t)) {
    $item_list = $row_t['keyword'];
    $category_list = $row_t['category'];
    $link_list = "<a href=\"$row_t[link]>click for buy</a>";
    $counts_list = $row_t['count'];

    array_push($items, $item_list);
    array_push($categories, $category_list);
    array_push($links, $link_list);
    array_push($counts, $counts_list);
}

$data = [
    ['keyword', 'count'],
    [$items[0], (int)$counts[0]],
    [$items[1], (int)$counts[1]],
    [$items[2], (int)$counts[2]],
    [$items[3], (int)$counts[3]],
    [$items[4], (int)$counts[4]]
]
```

The status bar at the bottom indicates: 줄 319, 열 52, 공백: 4, UTF-8, CRLF, PHP, Go Live.

By connecting with mysql, we show the data (stored in mysql) in the web page

### 3. Description

*Web programming – back using php & javascript*



contents



contents



contents



contents



contents



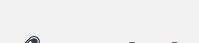
contents



contents



contents



contents

```
main.php -- php
main.php
90
91 <!DOCTYPE html>
92 <html lang="en">
93
94 <head>
95   <meta charset="UTF-8">
96   <meta name="viewport" content="width=device-width, initial-scale=1.0">
97   <title>main</title>
98   <link rel="stylesheet" href="css/navbar&footer.css">
99   <link rel="stylesheet" href="css/main.css">
100  <link rel="preconnect" href="https://fonts.gstatic.com">
101  <link href="https://fonts.googleapis.com/css2?family=Noto+Sans+Display&display=swap" rel="stylesheet">
102  <link rel="stylesheet" href="https://code.jquery.com/ui/1.11.2/themes/smoothness/jquery-ui.css">
103  <script src="http://code.jquery.com/jquery-1.10.2.js"></script>
104  <script src="https://code.jquery.com/ui/1.11.2/jquery-ui.js"></script>
105  <link rel="stylesheet" href="resource/demo/style.css">
106  <script src="https://code.jquery.com/jquery-3.2.1.min.js"></script>
107
108
109 <script src="https://www.google.com/jsapi"></script>
110 <script>
111   let data = <?= json_encode($data) ?>;
112   let options = <?= json_encode($options) ?>;
113   google.load('visualization', '1.0', {'packages':['corechart']});
114   google.setOnLoadCallback(function() {
115     let chart = new google.visualization.ColumnChart(document.querySelector('#chart_div'));
116     chart.draw(google.visualization.arrayToDataTable(data), options);
117   });
118 </script>
119
120 <style type="text/css">
121 .cal_top {
122   text-align: center;
123   font-size: 20px;
124 }
125
126 .cal {
127   text-align: center;
128 }
129
130 table.calendar {
131   border: 1px solid black;
132   display: inline-table;
133   text-align: left;
134 }
135

```

```
drama1.php -- php
drama1.php
92
93   <!-- word cloud -->
94   <script src="https://cdn.zingchart.com/zingchart.min.js"></script>
95
96   <html>
97     <body>
98       height: 100%;
99       width: 100%;
100      margin: 0;
101      padding: 0;
102
103      #myChart {
104        height: 100%;
105        width: 100%;
106        min-height: 150px;
107      }
108
109      .zc-ref {
110        display: none;
111      }
112    </style>
113
114    <head>
115      <title>TREN_CATCHER</title>
116    </head>
117
118    <nav class="upperbar">
119      <div class="logo">
120        <a href="main.php" style="text-decoration:none;">
121          <span class="upperbar_t1">TREN</span>
122          <span class="upperbar_t2">-D</span>
123          <span class="upperbar_t3">CATCHER</span>
124        </a>
125      </div>
126
127      <p class="welcome">
128        Welcome to TREN_CATCHER dashboard page!
129      </p>
130
131    </nav>
132
133    <main>
134      <div class="main_first">
135        <div class="main_first_icon">
136          <div class="main_profile">
137            <div class="main_profile_icon"></div>
138          </div>
139        </div>
140      </div>
141
142      <div class="main_second">
143        <div class="main_second_icon">
144          <div class="main_story">
145            <div class="main_story_content">
146              <div class="main_story_header">
147                <h2>TREN_CATCHER</h2>
148                <h3>Dashboard</h3>
149              </div>
150              <div class="main_story_main">
151                <div class="main_story_main_header">
152                  <h4>Top 10 Dramas</h4>
153                </div>
154                <div class="main_story_main_content">
155                  <table>
156                    <thead>
157                      <tr>
158                        <th>Rank</th>
159                        <th>Title</th>
160                        <th>Category</th>
161                        <th>Keywords</th>
162                      </tr>
163                    </thead>
164                    <tbody>
165                      <tr>
166                        <td>1</td>
167                        <td>Drama</td>
168                        <td>TV</td>
169                        <td>Korean, Romance, Thriller</td>
170                      </tr>
171                      <tr>
172                        <td>2</td>
173                        <td>Drama</td>
174                        <td>TV</td>
175                        <td>Korean, Romance, Thriller</td>
176                      </tr>
177                      <tr>
178                        <td>3</td>
179                        <td>Drama</td>
180                        <td>TV</td>
181                        <td>Korean, Romance, Thriller</td>
182                      </tr>
183                      <tr>
184                        <td>4</td>
185                        <td>Drama</td>
186                        <td>TV</td>
187                        <td>Korean, Romance, Thriller</td>
188                      </tr>
189                      <tr>
190                        <td>5</td>
191                        <td>Drama</td>
192                        <td>TV</td>
193                        <td>Korean, Romance, Thriller</td>
194                      </tr>
195                      <tr>
196                        <td>6</td>
197                        <td>Drama</td>
198                        <td>TV</td>
199                        <td>Korean, Romance, Thriller</td>
200                      </tr>
201                      <tr>
202                        <td>7</td>
203                        <td>Drama</td>
204                        <td>TV</td>
205                        <td>Korean, Romance, Thriller</td>
206                      </tr>
207                      <tr>
208                        <td>8</td>
209                        <td>Drama</td>
210                        <td>TV</td>
211                        <td>Korean, Romance, Thriller</td>
212                      </tr>
213                      <tr>
214                        <td>9</td>
215                        <td>Drama</td>
216                        <td>TV</td>
217                        <td>Korean, Romance, Thriller</td>
218                      </tr>
219                      <tr>
220                        <td>10</td>
221                        <td>Drama</td>
222                        <td>TV</td>
223                        <td>Korean, Romance, Thriller</td>
224                      </tr>
225                    </tbody>
226                  </table>
227                </div>
228              </div>
229            </div>
230          </div>
231        </div>
232      </div>
233
234      <div class="main_second">
235        <div class="main_second_icon">
236          <div class="main_story">
237            <div class="main_story_content">
238              <div class="main_story_header">
239                <h2>Word Cloud</h2>
240              </div>
241              <div class="main_story_main">
242                <div class="main_story_main_header">
243                  <h4>Top 10 Keywords</h4>
244                </div>
245                <div class="main_story_main_content">
246                  <img alt="Word cloud visualization showing the most frequent keywords from the dramas." data-bbox="100px 100px"/>
247                </div>
248              </div>
249            </div>
250          </div>
251        </div>
252      </div>
253
254    </main>
255
256    <div class="main_footer">
257      <div class="main_footer_content">
258        <div class="main_footer_left">
259          <div class="main_footer_left_header">
260            <h4>About Us</h4>
261          </div>
262          <div class="main_footer_left_content">
263            <p>TREN_CATCHER is a web-based dashboard application designed to analyze and visualize data from various Korean dramas. It provides users with insights into the most popular categories, genres, and keywords used in these dramas, allowing them to gain a deeper understanding of the trends and patterns within the Korean entertainment industry. The application uses advanced data processing and visualization techniques to make complex information accessible and engaging for users of all levels. Whether you're a drama enthusiast or a professional in the industry, TREN_CATCHER is the perfect tool for you!</p>
264          </div>
265        </div>
266        <div class="main_footer_right">
267          <div class="main_footer_right_header">
268            <h4>Contact Us</h4>
269          </div>
270          <div class="main_footer_right_content">
271            <p>Email: support@trendcatcher.com<br/>Phone: +123-456-7890<br/>Address: 123 Main Street, Anytown, USA</p>
272          </div>
273        </div>
274      </div>
275    </div>
276
277    <div class="main_footer">
278      <div class="main_footer_content">
279        <div class="main_footer_left">
280          <div class="main_footer_left_header">
281            <h4>Helpful Links</h4>
282          </div>
283          <div class="main_footer_left_content">
284            <ul>
285              <li>Home</li>
286              <li>Dashboard</li>
287              <li>Data Analysis</li>
288              <li>Visualizations</li>
289              <li>About Us</li>
290              <li>Contact Us</li>
291            </ul>
292          </div>
293        </div>
294        <div class="main_footer_right">
295          <div class="main_footer_right_header">
296            <h4>Follow Us</h4>
297          </div>
298          <div class="main_footer_right_content">
299            <ul>
300              <li>Facebook</li>
301              <li>Twitter</li>
302              <li>Instagram</li>
303              <li>YouTube</li>
304            </ul>
305          </div>
306        </div>
307      </div>
308    </div>
309
310    <div class="main_footer">
311      <div class="main_footer_content">
312        <div class="main_footer_left">
313          <div class="main_footer_left_header">
314            <h4>Feedback</h4>
315          </div>
316          <div class="main_footer_left_content">
317            <form>
318              <input type="text" placeholder="Your Name" data-bbox="100px 100px"/>
319              <input type="text" placeholder="Email Address" data-bbox="100px 100px"/>
320              <input type="text" placeholder="Feedback Message" data-bbox="100px 100px"/>
321              <input type="button" value="Submit" data-bbox="100px 100px"/>
322            </form>
323          </div>
324        </div>
325        <div class="main_footer_right">
326          <div class="main_footer_right_header">
327            <h4>Support</h4>
328          </div>
329          <div class="main_footer_right_content">
330            <ul>
331              <li>FAQ</li>
332              <li>Help Center</li>
333              <li>Contact Support</li>
334            </ul>
335          </div>
336        </div>
337      </div>
338    </div>
339
340    <div class="main_footer">
341      <div class="main_footer_content">
342        <div class="main_footer_left">
343          <div class="main_footer_left_header">
344            <h4>Legal</h4>
345          </div>
346          <div class="main_footer_left_content">
347            <ul>
348              <li>Terms of Service</li>
349              <li>Privacy Policy</li>
350              <li>Cookie Policy</li>
351              <li>Refund Policy</li>
352            </ul>
353          </div>
354        </div>
355        <div class="main_footer_right">
356          <div class="main_footer_right_header">
357            <h4>Documentation</h4>
358          </div>
359          <div class="main_footer_right_content">
360            <ul>
361              <li>API Reference</li>
362              <li>SDK Documentation</li>
363              <li>Sample Code</li>
364            </ul>
365          </div>
366        </div>
367      </div>
368    </div>
369
370    <div class="main_footer">
371      <div class="main_footer_content">
372        <div class="main_footer_left">
373          <div class="main_footer_left_header">
374            <h4>Community</h4>
375          </div>
376          <div class="main_footer_left_content">
377            <ul>
378              <li>User Forum</li>
379              <li>Developer Lounge</li>
380              <li>GitHub Repository</li>
381            </ul>
382          </div>
383        </div>
384        <div class="main_footer_right">
385          <div class="main_footer_right_header">
386            <h4>Partners</h4>
387          </div>
388          <div class="main_footer_right_content">
389            <ul>
390              <li>ABC Company</li>
391              <li>XYZ Corporation</li>
392              <li>PQR Inc.</li>
393            </ul>
394          </div>
395        </div>
396      </div>
397    </div>
398
399    <div class="main_footer">
400      <div class="main_footer_content">
401        <div class="main_footer_left">
402          <div class="main_footer_left_header">
403            <h4>Resources</h4>
404          </div>
405          <div class="main_footer_left_content">
406            <ul>
407              <li>Blog Posts</li>
408              <li>Case Studies</li>
409              <li>White Papers</li>
410            </ul>
411          </div>
412        </div>
413        <div class="main_footer_right">
414          <div class="main_footer_right_header">
415            <h4>Support</h4>
416          </div>
417          <div class="main_footer_right_content">
418            <ul>
419              <li>FAQ</li>
420              <li>Help Center</li>
421              <li>Contact Support</li>
422            </ul>
423          </div>
424        </div>
425      </div>
426    </div>
427
428    <div class="main_footer">
429      <div class="main_footer_content">
430        <div class="main_footer_left">
431          <div class="main_footer_left_header">
432            <h4>About Us</h4>
433          </div>
434          <div class="main_footer_left_content">
435            <ul>
436              <li>Our Story</li>
437              <li>Our Team</li>
438              <li>Our Mission</li>
439            </ul>
440          </div>
441        </div>
442        <div class="main_footer_right">
443          <div class="main_footer_right_header">
444            <h4>Partners</h4>
445          </div>
446          <div class="main_footer_right_content">
447            <ul>
448              <li>ABC Company</li>
449              <li>XYZ Corporation</li>
450              <li>PQR Inc.</li>
451            </ul>
452          </div>
453        </div>
454      </div>
455    </div>
456
457    <div class="main_footer">
458      <div class="main_footer_content">
459        <div class="main_footer_left">
460          <div class="main_footer_left_header">
461            <h4>Support</h4>
462          </div>
463          <div class="main_footer_left_content">
464            <ul>
465              <li>FAQ</li>
466              <li>Help Center</li>
467              <li>Contact Support</li>
468            </ul>
469          </div>
470        </div>
471        <div class="main_footer_right">
472          <div class="main_footer_right_header">
473            <h4>About Us</h4>
474          </div>
475          <div class="main_footer_right_content">
476            <ul>
477              <li>Our Story</li>
478              <li>Our Team</li>
479              <li>Our Mission</li>
480            </ul>
481          </div>
482        </div>
483      </div>
484    </div>
485
486    <div class="main_footer">
487      <div class="main_footer_content">
488        <div class="main_footer_left">
489          <div class="main_footer_left_header">
490            <h4>Partners</h4>
491          </div>
492          <div class="main_footer_left_content">
493            <ul>
494              <li>ABC Company</li>
495              <li>XYZ Corporation</li>
496              <li>PQR Inc.</li>
497            </ul>
498          </div>
499        </div>
500        <div class="main_footer_right">
501          <div class="main_footer_right_header">
502            <h4>Support</h4>
503          </div>
504          <div class="main_footer_right_content">
505            <ul>
506              <li>FAQ</li>
507              <li>Help Center</li>
508              <li>Contact Support</li>
509            </ul>
510          </div>
511        </div>
512      </div>
513    </div>
514
515    <div class="main_footer">
516      <div class="main_footer_content">
517        <div class="main_footer_left">
518          <div class="main_footer_left_header">
519            <h4>About Us</h4>
520          </div>
521          <div class="main_footer_left_content">
522            <ul>
523              <li>Our Story</li>
524              <li>Our Team</li>
525              <li>Our Mission</li>
526            </ul>
527          </div>
528        </div>
529        <div class="main_footer_right">
530          <div class="main_footer_right_header">
531            <h4>Partners</h4>
532          </div>
533          <div class="main_footer_right_content">
534            <ul>
535              <li>ABC Company</li>
536              <li>XYZ Corporation</li>
537              <li>PQR Inc.</li>
538            </ul>
539          </div>
540        </div>
541      </div>
542    </div>
543
544    <div class="main_footer">
545      <div class="main_footer_content">
546        <div class="main_footer_left">
547          <div class="main_footer_left_header">
548            <h4>Support</h4>
549          </div>
550          <div class="main_footer_left_content">
551            <ul>
552              <li>FAQ</li>
553              <li>Help Center</li>
554              <li>Contact Support</li>
555            </ul>
556          </div>
557        </div>
558        <div class="main_footer_right">
559          <div class="main_footer_right_header">
560            <h4>About Us</h4>
561          </div>
562          <div class="main_footer_right_content">
563            <ul>
564              <li>Our Story</li>
565              <li>Our Team</li>
566              <li>Our Mission</li>
567            </ul>
568          </div>
569        </div>
570      </div>
571    </div>
572
573    <div class="main_footer">
574      <div class="main_footer_content">
575        <div class="main_footer_left">
576          <div class="main_footer_left_header">
577            <h4>Partners</h4>
578          </div>
579          <div class="main_footer_left_content">
580            <ul>
581              <li>ABC Company</li>
582              <li>XYZ Corporation</li>
583              <li>PQR Inc.</li>
584            </ul>
585          </div>
586        </div>
587        <div class="main_footer_right">
588          <div class="main_footer_right_header">
589            <h4>Support</h4>
590          </div>
591          <div class="main_footer_right_content">
592            <ul>
593              <li>FAQ</li>
594              <li>Help Center</li>
595              <li>Contact Support</li>
596            </ul>
597          </div>
598        </div>
599      </div>
600    </div>
601
602    <div class="main_footer">
603      <div class="main_footer_content">
604        <div class="main_footer_left">
605          <div class="main_footer_left_header">
606            <h4>About Us</h4>
607          </div>
608          <div class="main_footer_left_content">
609            <ul>
610              <li>Our Story</li>
611              <li>Our Team</li>
612              <li>Our Mission</li>
613            </ul>
614          </div>
615        </div>
616        <div class="main_footer_right">
617          <div class="main_footer_right_header">
618            <h4>Partners</h4>
619          </div>
620          <div class="main_footer_right_content">
621            <ul>
622              <li>ABC Company</li>
623              <li>XYZ Corporation</li>
624              <li>PQR Inc.</li>
625            </ul>
626          </div>
627        </div>
628      </div>
629    </div>
630
631    <div class="main_footer">
632      <div class="main_footer_content">
633        <div class="main_footer_left">
634          <div class="main_footer_left_header">
635            <h4>Support</h4>
636          </div>
637          <div class="main_footer_left_content">
638            <ul>
639              <li>FAQ</li>
640              <li>Help Center</li>
641              <li>Contact Support</li>
642            </ul>
643          </div>
644        </div>
645        <div class="main_footer_right">
646          <div class="main_footer_right_header">
647            <h4>About Us</h4>
648          </div>
649          <div class="main_footer_right_content">
650            <ul>
651              <li>Our Story</li>
652              <li>Our Team</li>
653              <li>Our Mission</li>
654            </ul>
655          </div>
656        </div>
657      </div>
658    </div>
659
660    <div class="main_footer">
661      <div class="main_footer_content">
662        <div class="main_footer_left">
663          <div class="main_footer_left_header">
664            <h4>Partners</h4>
665          </div>
666          <div class="main_footer_left_content">
667            <ul>
668              <li>ABC Company</li>
669              <li>XYZ Corporation</li>
670              <li>PQR Inc.</li>
671            </ul>
672          </div>
673        </div>
674        <div class="main_footer_right">
675          <div class="main_footer_right_header">
676            <h4>Support</h4>
677          </div>
678          <div class="main_footer_right_content">
679            <ul>
680              <li>FAQ</li>
681              <li>Help Center</li>
682              <li>Contact Support</li>
683            </ul>
684          </div>
685        </div>
686      </div>
687    </div>
688
689    <div class="main_footer">
690      <div class="main_footer_content">
691        <div class="main_footer_left">
692          <div class="main_footer_left_header">
693            <h4>About Us</h4>
694          </div>
695          <div class="main_footer_left_content">
696            <ul>
697              <li>Our Story</li>
698              <li>Our Team</li>
699              <li>Our Mission</li>
700            </ul>
701          </div>
702        </div>
703        <div class="main_footer_right">
704          <div class="main_footer_right_header">
705            <h4>Partners</h4>
706          </div>
707          <div class="main_footer_right_content">
708            <ul>
709              <li>ABC Company</li>
710              <li>XYZ Corporation</li>
711              <li>PQR Inc.</li>
712            </ul>
713          </div>
714        </div>
715      </div>
716    </div>
717
718    <div class="main_footer">
719      <div class="main_footer_content">
720        <div class="main_footer_left">
721          <div class="main_footer_left_header">
722            <h4>Support</h4>
723          </div>
724          <div class="main_footer_left_content">
725            <ul>
726              <li>FAQ</li>
727              <li>Help Center</li>
728              <li>Contact Support</li>
729            </ul>
730          </div>
731        </div>
732        <div class="main_footer_right">
733          <div class="main_footer_right_header">
734            <h4>About Us</h4>
735          </div>
736          <div class="main_footer_right_content">
737            <ul>
738              <li>Our Story</li>
739              <li>Our Team</li>
740              <li>Our Mission</li>
741            </ul>
742          </div>
743        </div>
744      </div>
745    </div>
746
747    <div class="main_footer">
748      <div class="main_footer_content">
749        <div class="main_footer_left">
750          <div class="main_footer_left_header">
751            <h4>Partners</h4>
752          </div>
753          <div class="main_footer_left_content">
754            <ul>
755              <li>ABC Company</li>
756              <li>XYZ Corporation</li>
757              <li>PQR Inc.</li>
758            </ul>
759          </div>
760        </div>
761        <div class="main_footer_right">
762          <div class="main_footer_right_header">
763            <h4>Support</h4>
764          </div>
765          <div class="main_footer_right_content">
766            <ul>
767              <li>FAQ</li>
768              <li>Help Center</li>
769              <li>Contact Support</li>
770            </ul>
771          </div>
772        </div>
773      </div>
774    </div>
775
776    <div class="main_footer">
777      <div class="main_footer_content">
778        <div class="main_footer_left">
779          <div class="main_footer_left_header">
780            <h4>About Us</h4>
781          </div>
782          <div class="main_footer_left_content">
783            <ul>
784              <li>Our Story</li>
785              <li>Our Team</li>
786              <li>
```

## 4. Final result

Web page



contents



contents



contents



contents



contents



contents



contents



contents



contents



<http://3.35.218.159/trend/main.php>

## 5. Limitation



contents



contents

1. *Flume or oozie could not be used, and crontab was used to automate at a specified time.*



contents

-> *using flume or oozie to automate it*



contents



contents

2. *It tried to classify industrial groups using 'Naver shopping', but due to the characteristic of Twitter, keywords are not completely classified as industrial groups.*



contents

-> *link that comes out when you search for Naver shopping*



contents

3. *Cannot use spark*