ddit-scraping-for-week6-project-1

April 17, 2024

Be sure to install the `praw` and `wordcloud` packages if you don't have them. Using Anaconda and `conda install praw wordcloud` works best, but you can use `pip` too. The next cell will install `praw` via conda (the ! runs the command on the command line).

[1]:
```
!conda install -c conda-forge praw wordcloud -y
```

```
Collecting package metadata (current_repodata.json): …working… done
Solving environment: …working… done

# All requested packages already installed.



==> WARNING: A newer version of conda exists. <==
  current version: 23.7.4
  latest version: 24.3.0

Please update conda by running

    $ conda update -n base -c defaults conda

Or to minimize the number of packages updated during conda update use

     conda install conda=24.3.0
```

[2]:
```
import praw
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

Create a Reddit account, then go to https://www.reddit.com/prefs/apps .

This subreddit is here: https://www.reddit.com/r/socialjustice101/top/?t=all

[19]:
```
import praw

# Ensure these variables are correctly assigned with your credentials
```

```
client_id = 'HHVfGcajhHPaTSh4IqVbDA'
client_secret = 'ZR5tjjEBtABsO6tYdLnQFwgBLSOGyg'
user_agent = 'ricky884213'

# Initialize the Reddit instance
reddit = praw.Reddit(client_id=client_id,
                     client_secret=client_secret,
                     user_agent=user_agent)

try:
    # Accessing the subreddit and getting top posts
    subreddit = reddit.subreddit('BasicIncome').top(time_filter='all',␣
 ↪limit=100)

    # Dictionary to store post titles and bodies
    comments_dict = {'title': [], 'body': []}

    # Iterate through top posts and collect data
    for post in subreddit:
        comments_dict['title'].append(post.title)
        comments_dict['body'].append(post.selftext)

except Exception as e:
    print(f"An error occurred: {e}")
```

[20]:
```
wc = WordCloud().generate(' '.join(comments_dict['body']))
plt.imshow(wc, interpolation="bilinear")
plt.axis('off')
```

[20]: (-0.5, 399.5, 199.5, -0.5)

[21]:
```
import praw

# Ensure these variables are correctly assigned with your credentials
client_id = 'HHVfGcajhHPaTSh4IqVbDA'
client_secret = 'ZR5tjjEBtABsO6tYdLnQFwgBLSOGyg'
user_agent = 'ricky884213'

# Initialize the Reddit instance
reddit = praw.Reddit(client_id=client_id,
                     client_secret=client_secret,
                     user_agent=user_agent)

try:
    # Accessing the subreddit and getting top posts
    subreddit = reddit.subreddit('incomeinequality').top(time_filter='all',␣
 ↪limit=100)

    # Dictionary to store post titles and bodies
    comments_dict = {'title': [], 'body': []}

    # Iterate through top posts and collect data
    for post in subreddit:
        comments_dict['title'].append(post.title)
        comments_dict['body'].append(post.selftext)

except Exception as e:
    print(f"An error occurred: {e}")
```

[22]:
```
wc = WordCloud().generate(' '.join(comments_dict['body']))
plt.imshow(wc, interpolation="bilinear")
plt.axis('off')
```

[22]:
```
(-0.5, 399.5, 199.5, -0.5)
```

References :

1.Python Reddit API Wrapper (PRAW)
2.https://www.reddit.com/r
3.reffered - Week 7 – Collecting Social Media Data (MSDS600_S41_Introduction to Data Science)

[ ]: