



Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

Лабораторна робота №3
з дисципліни
«Web - аналітика»

Виконав:
студент групи ФБ-31мп
Щур Павло
Перевірив:
Ткач В. М.

Київ-2024

Посилання на GitHub: <https://github.com/ShchurPavlo/web-analytics-2024/tree/main/lab3>

Завдання

1. Використовуючи API Twitter, написати скрипт, що буде автентифікуватися в соціальній мережі та забирати $N > 1000$ останніх записів певного акаунту (акаунт має бути публічним), та створити датасет. Опціонально – сформувати систему сповіщення (пошта, телефон, будь-що) про знаходження певного ключового слова.

Виконання:

1) Замість платного API Twitter будемо використовувати альтернативу – API Reddit. Зареєструємося там та отримаємо ключі доступу до API:

developed applications



The screenshot shows the 'developed applications' page for a user named 'Pavlo Shchur FB-31mp'. The application is named 'Lab3_Shchur_App' and is a 'web app'. It has a description 'Pavlo Shchur FB-31mp' and a 'redirect uri' of 'https://webanalytics.com.ua'. The page includes fields for 'secret', 'name', 'description', 'about url', and 'redirect uri'. There is an 'update app' button and a list of developers, currently showing 'Impossible Seesaw483 (that's you!)' with a 'remove' link. There is also an 'add developer:' field and a 'delete app' link at the bottom.

2) Наступним кроком реалізуємо функцію автентифікації через API:

```
7 def Authorization():
8     id = input("Enter client id: ")
9     secret = input("Enter client secret: ")
10    user_agent = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/122.0.0.0 Safari/537.36'
11    try:
12        client=praw.Reddit(client_id=id,client_secret=secret,user_agent=user_agent)
13        print("Authentication successful!")
14    except Exception as e:
15        print("Authentication failed:", e)
16        exit()
17    return client
```

3) Також реалізуємо функцію сканування аккаунту та зберігання інформації про пости в датафрейм pandas:

```
21 def Parse(client,target):
22     result = []
23     for submission in client.redditor(target).submissions.new(limit=2000):
24         if submission.is_self and submission.selftext:
25             post_text = submission.selftext
26         else:
27             post_text = None
28             result.append({
29                 'id': submission.id,
30                 'author': submission.author,
31                 'title': submission.title,
32                 'number_of_comments': submission.num_comments,
33                 'created_time': datetime.fromtimestamp(submission.created_utc, tz=timezone.utc),
34                 'post_text': post_text
35             })
36     return pd.DataFrame(result)
```

Будемо збирати наступну інформацію:

- Ідентифікатор допису
- Автор
- Заголовок
- Кількість коментарів
- Час створення
- Текстовий зміст посту

4) В результаті отримаємо наступну інформацію:

```
Authentication successful!
   id  author  ...  created_time  post_text
0  3dwp71  ProxMox  ...  2015-07-20 04:03:08+00:00  None
1  2mcqk6  ProxMox  ...  2014-11-15 04:33:45+00:00  None
2  2ipt7w  ProxMox  ...  2014-10-09 00:58:32+00:00  None
3  2ep40d  ProxMox  ...  2014-08-27 04:31:25+00:00  None
4  2e59yk  ProxMox  ...  2014-08-21 03:10:00+00:00  None
5  2e58n5  ProxMox  ...  2014-08-21 02:55:41+00:00  None
6  27vyn1  ProxMox  ...  2014-06-11 16:32:17+00:00  None
7  1vm74w  ProxMox  ...  2014-01-19 20:48:06+00:00  None
8  1si5qc  ProxMox  ...  2013-12-09 23:58:05+00:00  None
9  1kduq8  ProxMox  ...  2013-08-14 23:46:02+00:00  None
10 1irzyc  ProxMox  ...  2013-07-21 23:27:48+00:00  None
11 1gl623  ProxMox  ...  2013-06-18 15:06:11+00:00  None
```

```
41 print(result_df)
42 result_df.to_csv( path_or_buf: 'Proxmox_df.csv', index=False)
```

