

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

WEB-Аналітика

ЛАБОРАТОРНА РОБОТА №3

Виконав

студент групи ФІ-42мн

Беш Радомир Андрійович

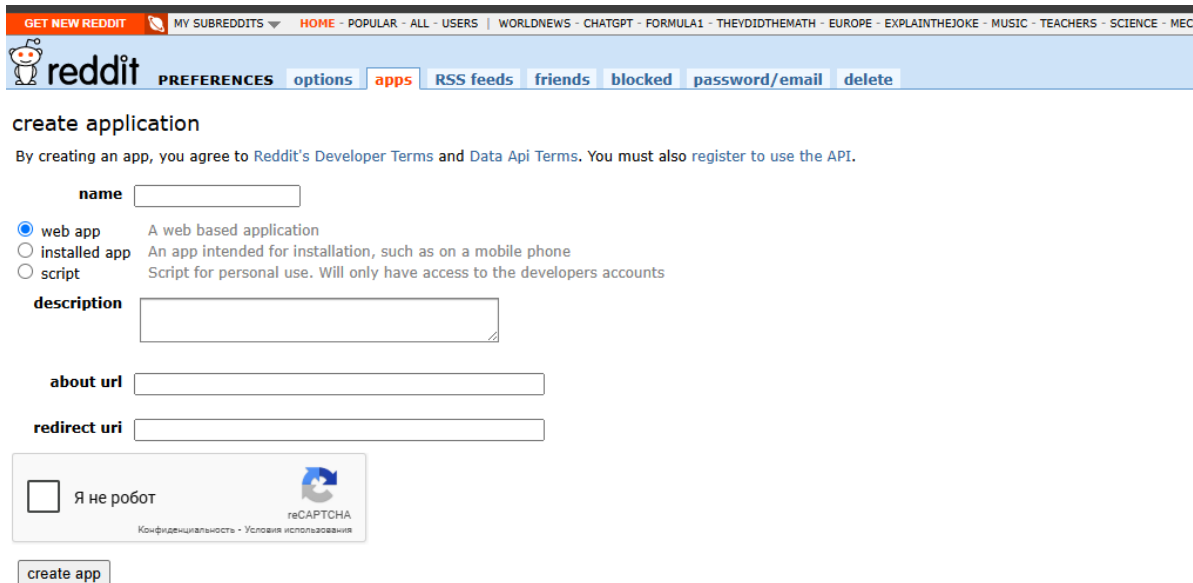
Київ 2025

Завдання:

Зробити парсинг певної соцмережі Reddit та створити БД для виявлення певних повідомлень (ключових слів). Зробити аналіз на виявлення аномальної поведінки(наприклад використання тригерних слів).

Хід роботи

- 1) Спочатку створимо користувача (акаунт) та додаток, щоб підключитися до Reddit API.



GET NEW REDDIT MY SUBREDDITS HOME - POPULAR - ALL - USERS WORLDNEWS - CHATGPT - FORMULA1 - THEYDIDTHEMATH - EUROPE - EXPLAINTHEJOKE - MUSIC - TEACHERS - SCIENCE - MECHANICS

reddit PREFERENCES options apps RSS feeds friends blocked password/email delete

create application

By creating an app, you agree to Reddit's Developer Terms and Data Api Terms. You must also register to use the API.

name

☒ web app A web based application
☐ installed app An app intended for installation, such as on a mobile phone
☐ script Script for personal use. Will only have access to the developers accounts

description

about url

redirect uri

☐ Я не робот reCAPTCHA
Конфиденциальность - Условия использования

create app

create application

By creating an app, you agree to Reddit's Developer Terms and Data Api Terms. You must also register to use the API.

name

☐ web app A web based application
☐ installed app An app intended for installation, such as on a mobile phone
☒ script Script for personal use. Will only have access to the developers accounts

description


about url

redirect uri a url is required

☒ Я не робот reCAPTCHA
Конфиденциальность - Условия использования

create app

developed applications

**lab_3** data reseach

change icon

secret [redacted]

name

description

about url

redirect uri

developers Aggressive_Issue5530 (that's you!) [remove](#)

add developer:

[delete app](#)

Отримали credentials для під'єднання.

2) Далі працюємо в Google Collab

Вводимо наші credentials та імпортуємо бібліотеки для роботи

```
import praw
import time
from datetime import datetime
import sqlite3

# Авторизація через Reddit API
reddit = praw.Reddit(
    client_id="[redacted]",
    client_secret="[redacted]",
    user_agent="lab_3 by /u/Aggressive_Issue5530"
)
```

Виконуємо пошук останніх постів за такою темою

```
subreddit = reddit.subreddit("ukraine")
posts = list(subreddit.new(limit=1000))
```

Створюємо БД і в ній таблицю для збереження даних, в нашому випадку це SQLite

```
conn = sqlite3.connect("reddit_posts2.db")
cursor = conn.cursor()
|
cursor.execute("""
    CREATE TABLE IF NOT EXISTS posts (
        id TEXT PRIMARY KEY,
        title TEXT,
        author TEXT,
        date TEXT,
        score INTEGER,
        num_comments INTEGER,
        url TEXT,
        keyword TEXT
    )
""")

conn.commit()
```

Перевіряємо створену таблицю

```
cursor.execute("PRAGMA table_info(posts);")
columns = cursor.fetchall()
for col in columns:
    print(col)

(0, 'id', 'TEXT', 0, None, 1)
(1, 'title', 'TEXT', 0, None, 0)
(2, 'author', 'TEXT', 0, None, 0)
(3, 'date', 'TEXT', 0, None, 0)
(4, 'score', 'INTEGER', 0, None, 0)
(5, 'num_comments', 'INTEGER', 0, None, 0)
(6, 'url', 'TEXT', 0, None, 0)
(7, 'keyword', 'TEXT', 0, None, 0)
```

Заносимо всі дані в таблицю

```
keyword = "war"
for post in posts:
    post_id = post.id
    title = post.title
    author = str(post.author)
    created = datetime.utcfromtimestamp(post.created_utc).strftime('%Y-%m-%d %H:%M:%S')
    score = post.score
    comments = post.num_comments
    url = post.url

    try:
        cursor.execute("""
            INSERT INTO posts (id, title, author, date, score, num_comments, url, keyword)
            VALUES (?, ?, ?, ?, ?, ?, ?, ?)
            """, (post_id, title, author, created, score, comments, url, keyword))
    except Exception as e:
        print("Ошибка при вставке:", e)

conn.commit()
```

За допомогою пандасу прочитаємо файл

```
import pandas as pd

conn = sqlite3.connect('reddit_posts2.db')
query = "SELECT * FROM posts order by score desc"
df = pd.read_sql_query(query, conn)
df
```

	id	title	author	date	score	num_comments	url	keyword
0	1kfj5cw	When Zelensky was in Prague, he noticed Ukrain...	ownworldman	2025-05-05 18:33:56	17198	195	https://i.redd.it/xqeqw03c0ze1.jpeg	war
1	1kh1zar	Swampy one of the best EOD technicians in Ukra...	mateiescu	2025-05-07 16:40:03	13688	237	https://i.redd.it/dy32778f1eze1.jpeg	war
2	1kmajzm	Beware of the Russians who bring peace	A_Lazko	2025-05-14 09:04:00	11524	157	https://i.redd.it/0e4mbx1jqp0f1.jpeg	war
3	1kk3z7d	Pope Leo XIV, in his address today, calls for ...	vectorix108	2025-05-11 15:43:54	10158	115	https://v.redd.it/chlq53n7b60f1	war
4	1kfV20h	Daniel Kovalchuk died. Rest in peace hero.	HarakenQQ	2025-05-06 03:34:02	9999	107	https://i.redd.it/2etxr0zf03ze1.jpeg	war
...
973	1keme9e	War memorabilia/trophies in support	coilsfromthedeath	2025-05-04 15:24:49	8	3	https://www.reddit.com/r/ukraine/comments/1kem...	war
974	1ki1mvj	Ніколи не проходив медогляд, тож не маю військ...	ConsalderCordo	2025-05-08 21:22:08	7	3	https://www.reddit.com/r/ukraine/comments/1ki1...	war
975	1ki2wto	what is the song or the artist in this music?	MildlyBear	2025-05-12 20:22:04	3	0	https://youtu.be/-sRxvQcvIk?t=2232	war
976	1kqcspr	Trump to speak to Putin with goal of reaching ...	NilsuBerk	2025-05-19 14:16:08	0	13	https://www.bbc.com/news/live/cz9y5xy81jqt	war
977	1kh5hh6	Can I visit Odesa?	Mr_a_bit_silly	2025-05-07 18:59:47	0	44	https://www.reddit.com/r/ukraine/comments/1kh5...	war

978 rows x 8 columns

Пошук акаунтів, які запостили більше ніж 10 постів за останній час

```
from collections import Counter
from datetime import datetime, timedelta
import matplotlib.pyplot as plt

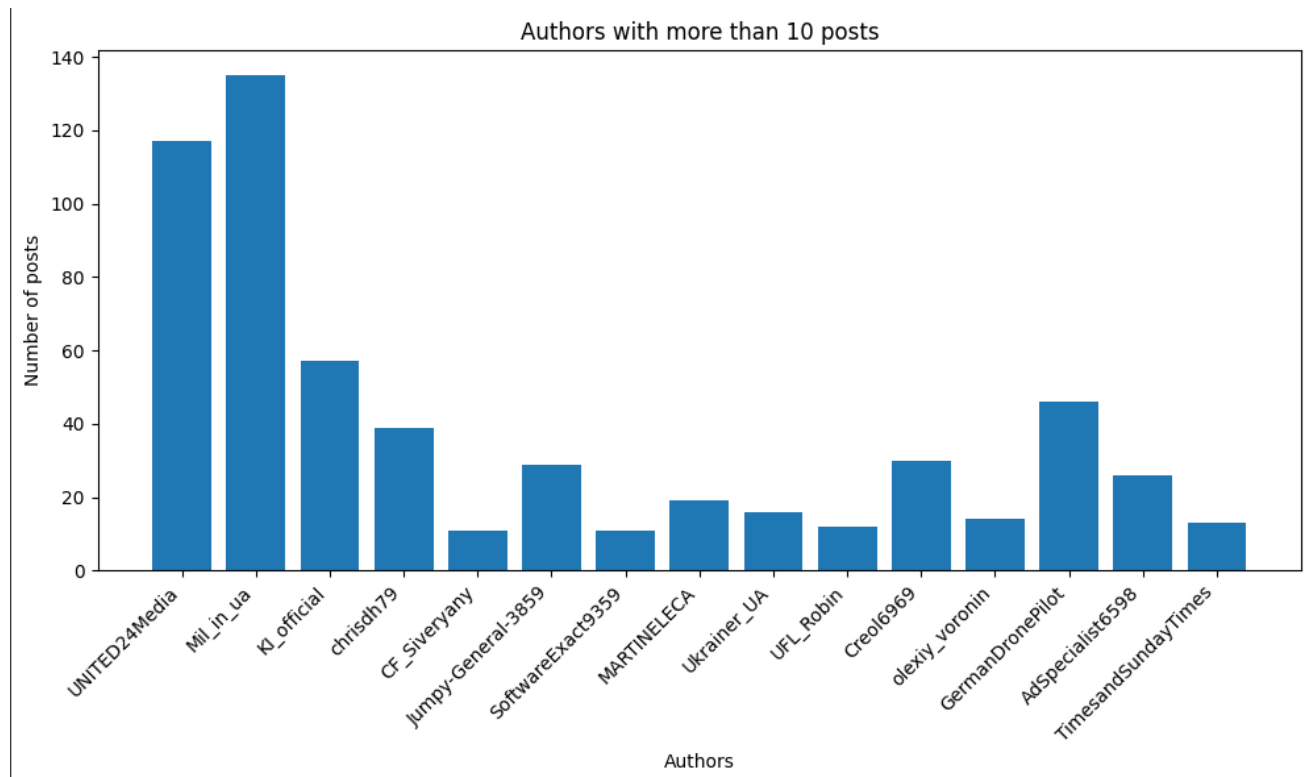
authors = []

for post in posts:
    authors.append(post.author.name if post.author else '[deleted]')

counter = Counter(authors)

for author, count in counter.items():
    if count > 10:
        print(f"{author}: {count} posts")

UNITED24Media: 117 posts
Mil_in_ua: 135 posts
KI_official: 57 posts
chrisdh79: 39 posts
CF_Siveryany: 11 posts
Jumpy-General-3859: 29 posts
SoftwareExact9359: 11 posts
MARTINELECA: 19 posts
Ukrainer_UA: 16 posts
UFL_Robin: 12 posts
Creol6969: 30 posts
olexiy_voronin: 14 posts
GermanDronePilot: 46 posts
AdSpecialist6598: 26 posts
TimesandSundayTimes: 13 posts
```



Спроба знайти по тригінним словам

```
trigger_words = [

    # Эмоциональные/резонансные слова
    "atrocitiy", "massacre", "genocide", "torture",
    "war crime", "crimes against humanity",
    "terror", "terrorist", "terrorism",
    "refugee", "displaced", "civilians killed", "children killed",

    # Военная терминология
    "counteroffensive", "advance", "retreat", "NATO", "air defense",
    "troops", "tank", "brigade", "mobilized", "draft", "conscription",
    "Kadyrov", "Wagner", "mercenaries",

    # Политически заряженные термины
    "Russia", "Putin", "Kremlin", "propaganda",
    "Zelenskyy", "Ukrainian army", "AFU",
    "denazification", "liberation",
    "occupied territories", "Donbas", "Crimea",

    # Дезинформация / паника
    "fake", "hoax", "false flag", "bio labs", "nuclear", "dirty bomb",
    "chemical weapons", "WW3", "world war"
]

mask = df["title"].apply(lambda x: any(word in x for word in trigger_words))

target_df = df[mask]

target_df[["author", "title", "url"]]
```

Маємо доволі багато постів:

	author	title	url
2	A_Lazko	Beware of the Russians who bring peace	https://i.redd.it/0e4mbx1jqp0f1.jpeg
5	UNITED24Media	Ukrainian mayor gave up his POW swap so his si...	https://i.redd.it/9z4lrso68c0f1.jpeg
6	PjeterPannos	Oleksandr Matsiyevsky — the heroic Ukrainian s...	https://i.redd.it/s5ip094uoyze1.jpeg
7	UNITED24Media	Zelenskyy sends a clear message to Putin from ...	https://v.redd.it/dfp7yfe36y0f1
8	TheGreatPornholio123	Putin unveils new massively useless weapon...	https://v.redd.it/gumr9ywtxd0f1
...
947	UNITED24Media	Pope Leo XIV Reaffirms Vatican's Readiness to ...	https://united24media.com/latest-news/pope-leo...
950	white1984	Putin loved this car. Now it drives Ukraine's ...	https://youtu.be/_sBuCJgM28E?si=EjIKO2k4kDLwt1_9
956	Lion8330	Murals with allusion to Makhno in Huliaipole, ...	https://www.youtube.com/watch?v=UslxO3seSqQ
961	Min_Min_Drops	What happens to the properties of civils in oc...	https://www.reddit.com/r/ukraine/comments/1kef...
976	NilsuBerk	Trump to speak to Putin with goal of reaching ...	https://www.bbc.com/news/live/cz9y5xy81jqt
487 rows × 3 columns			

Також спроба пошукати акаунти, які зробили 1 або 2 пости:

```
author_counts = target_df.groupby('author').size().reset_index(name='post_count')

# Фільтруємо авторів з 1 или 2 постами
low_activity_authors = author_counts[author_counts['post_count'].isin([1, 2])]
filtered_df = target_df[target_df['author'].isin(low_activity_authors['author'])]
filtered_df[['author', 'title', 'url']]
```

	author	title	url
6	PjeterPannos	Oleksandr Matsiyevsky — the heroic Ukrainian s...	https://i.redd.it/s5ip094uoyze1.jpeg
8	TheGreatPornholio123	Putin unveils new massively useless weapon...	https://v.redd.it/gumr9ywtxd0f1
16	Glavurdan	Trump: Russia will have to give up all of Ukraine	https://babel.ua/en/news/117703-trump-russia-w...
19	ijnfrt	A Russian soldier films the moment his group i...	https://v.redd.it/m4t5svcb3i0f1
27	Practical-Pea-1205	Majority of Americans support weapon delieveri...	https://kyivindependent.com/majority-of-ameri...
...
921	Due_Search_8040	Situation Report: Ceasefire Negotiations with ...	https://www.opforjournal.com/p/situation-repor...
939	Turborapt0r	What destroyed orc tanks are on display in Kyi...	https://www.reddit.com/r/ukraine/comments/1klh...
950	white1984	Putin loved this car. Now it drives Ukraine's ...	https://youtu.be/_sBuCJgM28E?si=EjiKO2k4kDLwt1_9
961	Min_Min_Drops	What happens to the properties of civils in oc...	https://www.reddit.com/r/ukraine/comments/1kef...
976	NilsuBerk	Trump to speak to Putin with goal of reaching ...	https://www.bbc.com/news/live/cz9y5xy81jq1

82 rows x 3 columns

Висновки: у ході виконання лабораторної роботи було здійснено повний цикл збору та обробки даних із соціальної мережі Reddit. Спочатку було створено обліковий запис та додаток для підключення до Reddit API, після чого здійснено підключення до API та зібрано пости за заданою тематикою. Отримані дані були збережені у створену базу даних на основі SQLite. За допомогою бібліотеки Pandas проведено попередній аналіз, зокрема виявлено акаунти, які публікували понад десять повідомлень, а також виконано пошук за тригерними словами, що потенційно вказують на аномальну поведінку. Але на жаль або на щастя аномалій і ботів не було знайдено. Проведений аналіз підтверджує можливість виявлення підозрілих чи нетипових дій користувачів, що може бути корисним для автоматизованого моніторингу контенту. Загалом, виконання цієї роботи сприяло поглибленню практичних знань у сфері роботи з API, обробки даних та виявлення аномалій у поведінці користувачів у соціальних мережах.

