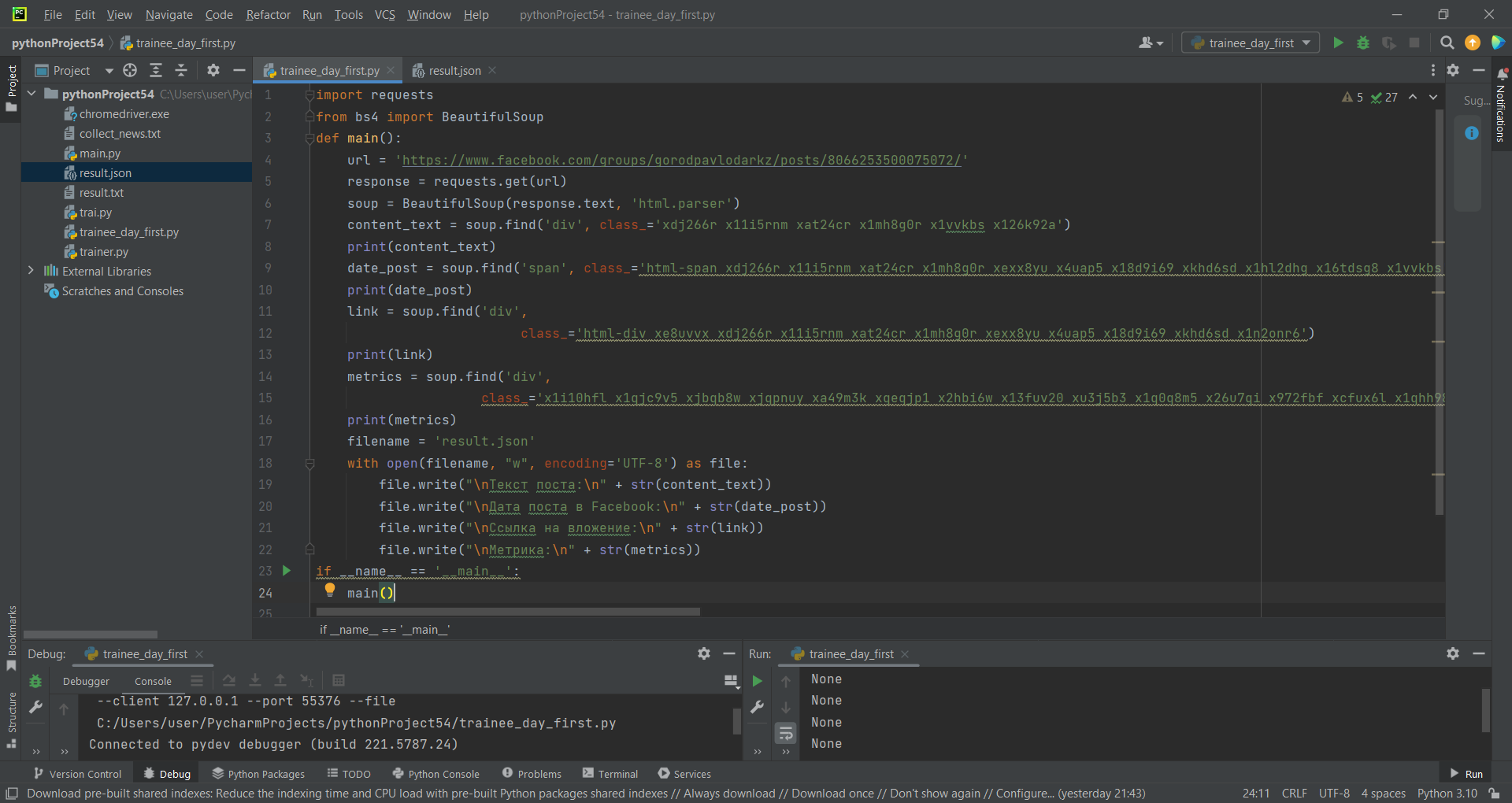
Решение Парсер Facebook

Как было рекомендовано "Попробуйте сперва парсить с помощью Requests и BeautifulSoup."

Попробовало. Не дал результатов. Не получилось получить "содержимого поста из Facebook-групп: текст, дату и ссылку на вложение". Ответ был везде None. (ссылка - https://www.facebook.com/groups/gorodpavlodarkz/posts/8066253500075072/ )



import requests

from bs4 import BeautifulSoup

def main():

url = 'https://www.facebook.com/groups/gorodpavlodarkz/posts/8066253500075072/'

response = requests.get(url)

soup = BeautifulSoup(response.text, 'html.parser')

content\_text = soup.find('div', class\_='xdj266r x11i5rnm xat24cr x1mh8g0r x1vvkbs x126k92a')

print(content\_text)

date\_post = soup.find('span', class\_='html-span xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkhd6sd x1hl2dhg x16tdsg8 x1vvkbs')

print(date\_post)

link = soup.find('div',

class\_='html-div xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r xexx8yu x4uap5 x18d9i69 xkhd6sd x1n2onr6')

print(link)

metrics = soup.find('div',

class\_='x1i10hfl x1qjc9v5 xjbqb8w xjqpnuy xa49m3k xqeqjp1 x2hbi6w x13fuv20 xu3j5b3 x1q0q8m5 x26u7qi x972fbf xcfux6l x1qhh985 xm0m39n x9f619 x1ypdohk xdl72j9 x2lah0s xe8uvvx xdj266r x11i5rnm xat24cr x1mh8g0r x2lwn1j xeuugli xexx8yu x4uap5 x18d9i69 xkhd6sd x1n2onr6 x16tdsg8 x1hl2dhg x1ja2u2z x1t137rt x1o1ewxj x3x9cwd x1e5q0jg x13rtm0m x3nfvp2 x1q0g3np x87ps6o x1lku1pv x1a2a7pz x5ve5x3')

print(metrics)

filename = 'result.json'

with open(filename, "w", encoding='UTF-8') as file:

file.write("\nТекст поста:\n" + str(content\_text))

file.write("\nДата поста в Facebook:\n" + str(date\_post))

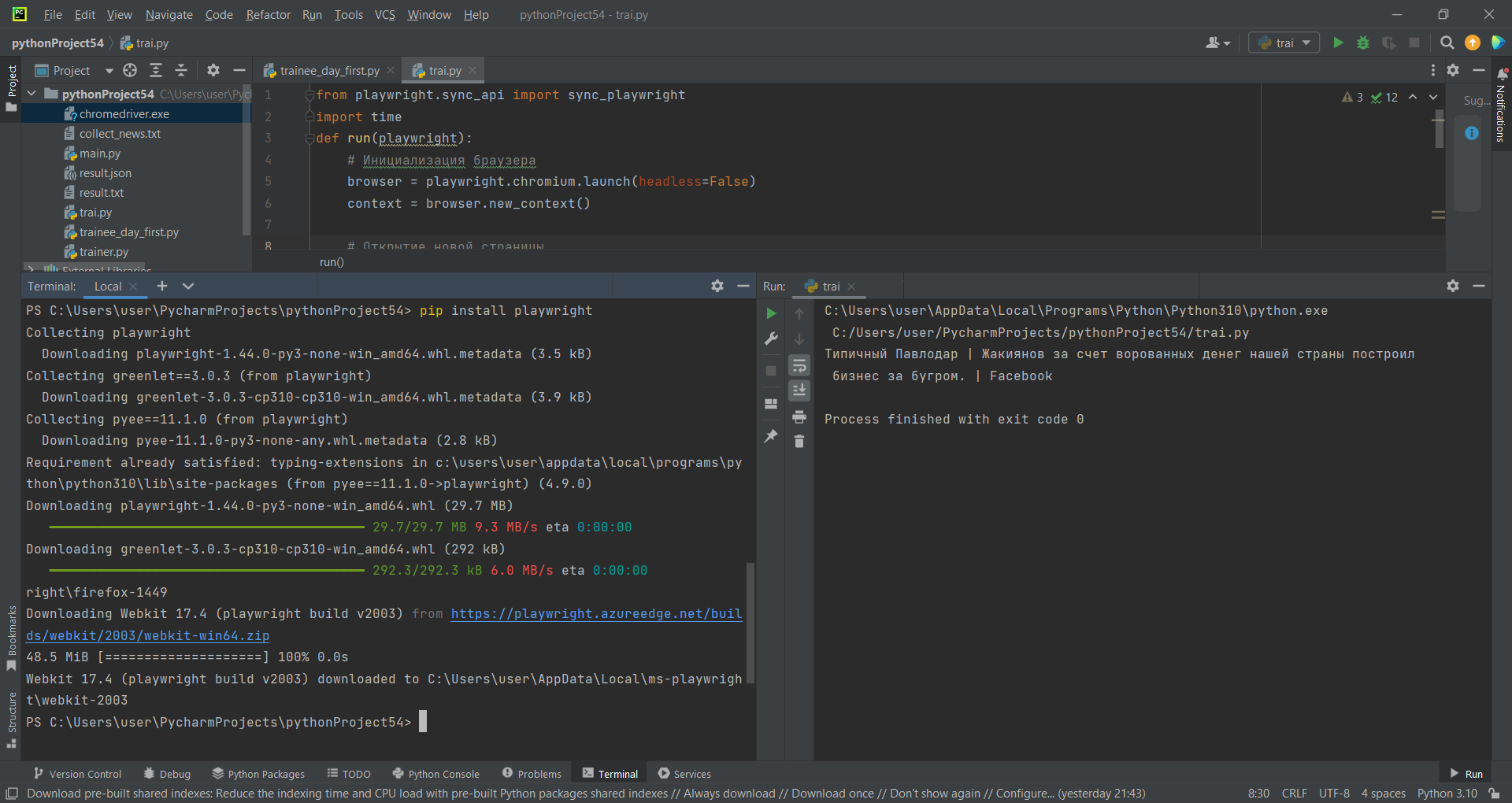
file.write("\nСсылка на вложение:\n" + str(link))

file.write("\nМетрика:\n" + str(metrics))

if \_\_name\_\_ == '\_\_main\_\_':

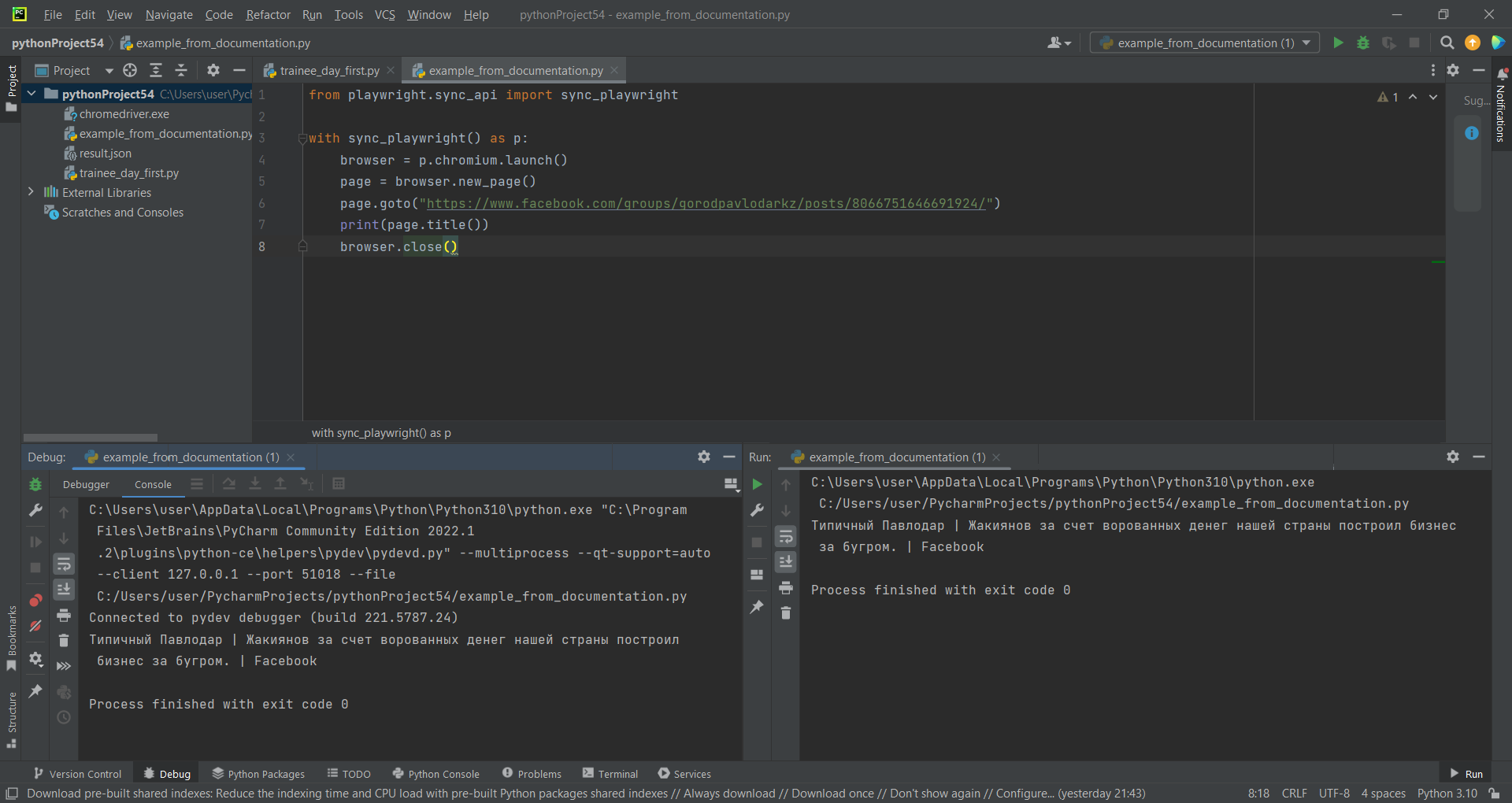
main()

Если это не даст результатов, попробуйте использовать инструменты автоматизации браузеров, такими как Selenium или Playwright.

Установка библиотеки Playwright

Читаем документацию и поэтапно пробуем использовать на практике

<https://playwright.dev/python/docs/library>



from playwright.sync\_api import sync\_playwright

with sync\_playwright() as p:

browser = p.chromium.launch()

page = browser.new\_page()

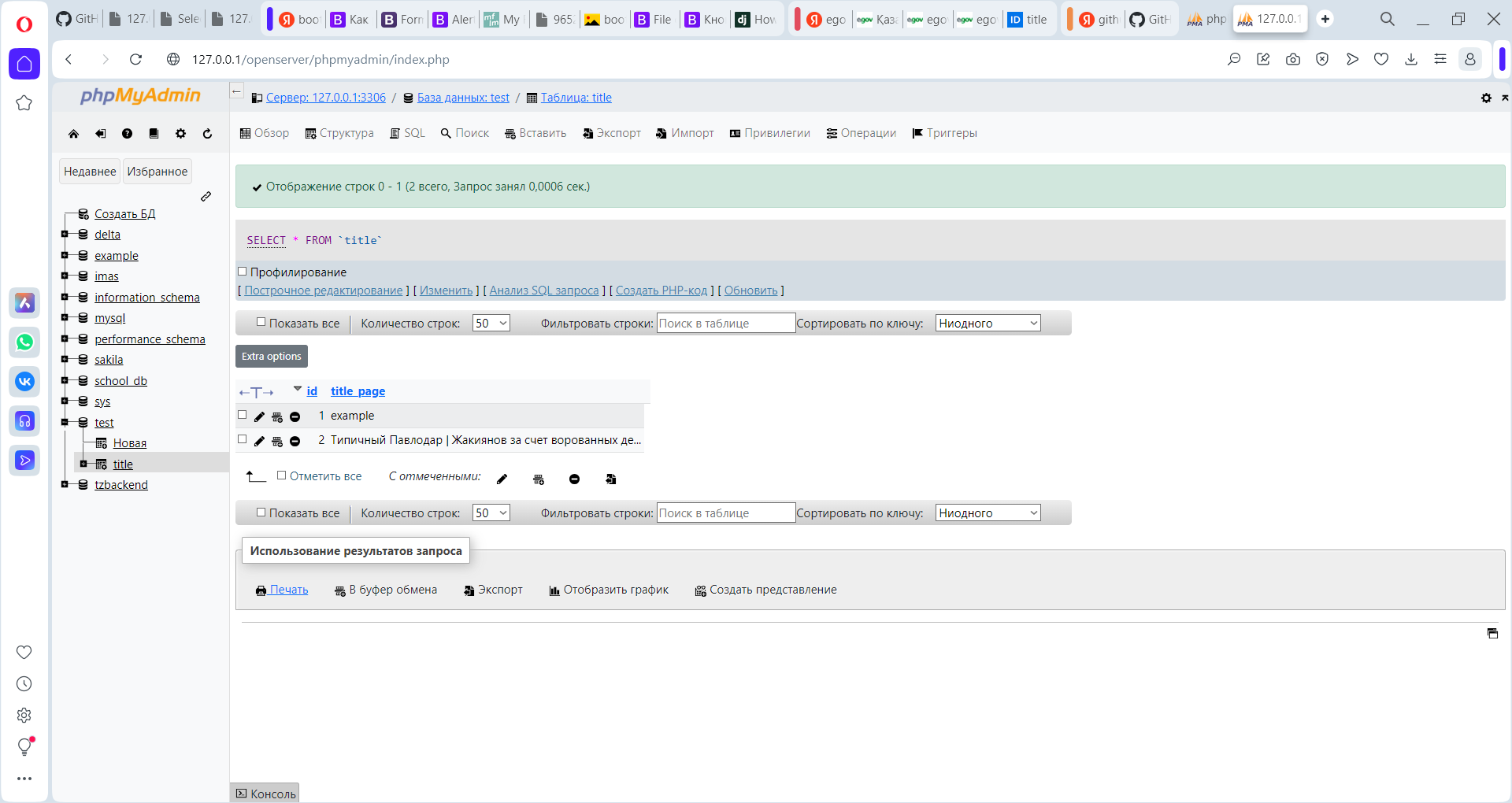
page.goto("https://www.facebook.com/groups/gorodpavlodarkz/posts/8066751646691924/")

print(page.title())

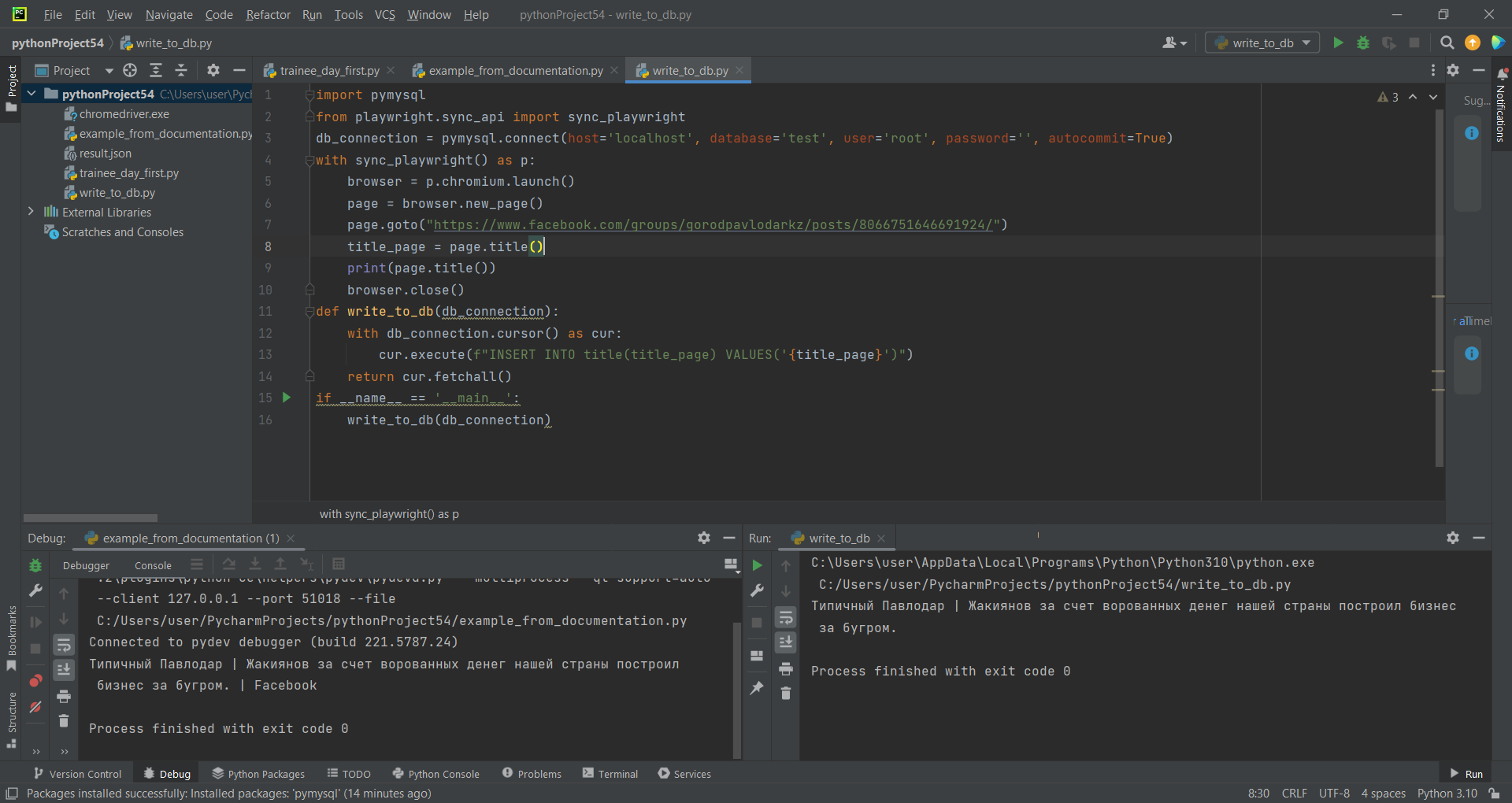
browser.close()

Код выше дал результат получение **title**. (это вообще правильно)

Создано база данных под названием **test** и таблица **title.**



Полученный title документа был записан в базу.



import pymysql

from playwright.sync\_api import sync\_playwright

db\_connection = pymysql.connect(host='localhost', database='test', user='root', password='', autocommit=True)

with sync\_playwright() as p:

browser = p.chromium.launch()

page = browser.new\_page()

page.goto("https://www.facebook.com/groups/gorodpavlodarkz/posts/8066751646691924/")

title\_page = page.title()

print(page.title())

browser.close()

def write\_to\_db(db\_connection):

with db\_connection.cursor() as cur:

cur.execute(f"INSERT INTO title(title\_page) VALUES('{title\_page}')")

return cur.fetchall()

if \_\_name\_\_ == '\_\_main\_\_':

write\_to\_db(db\_connection)

Отправляю то, что получилось: Это название **title** (правильно или нет) и записать **title**  в базу.

Что не вышло: Остальные части не получилось, потому что хочу найти через **class,** а ответ **None.** (Что делаю не так)