НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

Лабораторна робота №4 з дисципліни

«Програмування інтелектуальних інформаційних систем»

Прийняв

аспірант кафедри ІПІ

Баришич Л.М.

Виконав

студент групи ІП-12

Казаков В.О.

“26” листопада 2023 р.

Київ – 2023

Імпорт необхідних модулів:

from selenium import webdriver

from selenium.webdriver.common.by import By

import re

Зчитування імен з сайту:

driver = webdriver.Chrome()

driver.maximize\_window()

driver.get('https://pestrecy-rt.ru/news/tag/list/specoperaciia')

def get\_russian\_names(text):

russian\_names = re.findall(r'\b[А-ЯЁ][а-яё]+\s[А-ЯЁ][а-яё]+\b', text)

return russian\_names

name\_list = []

while True:

for headline in driver.find\_elements(By.XPATH, '//h2[@class="oneNews\_\_link news\_\_bold-text"]'):

names = get\_russian\_names(headline.text)

if names:

for name in names:

name\_list.append(name)

for paragraph in driver.find\_elements(By.XPATH, '//p[@class="oneNews\_\_link"]'):

names = get\_russian\_names(paragraph.text)

if names:

for name in names:

name\_list.append(name)

buttons = driver.find\_elements(By.XPATH, '//a[@class="all-news\_\_button\_forward"]//span[contains(text(), "Далее")]')

if buttons:

buttons[0].click()

else:

break

Фільтрація та виведення результатів:

banned\_substrings = ['Казан', 'Росси', 'Глав', 'Социальн', 'Урожен',

'Татарстан', 'Госдум', 'Вооружен', 'Защитник', 'Пестречин', 'Родин',

'Государст', 'Героя']

filtered\_name\_list = [item for item in name\_list if all(substring not

in item for substring in banned\_substrings)]

for name in filtered\_name\_list:

print(name)

