



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
ФАКУЛЬТЕТ СОЦИАЛЬНЫХ НАУК
ОП Политология

**"Искатель 3000": руководство
пользователя**

Выполнили:
Команда 14

1 О программе

2 Парсинг сайта Декларатор.орг

2.1 Установка Selenium

Для начала работы необходимо установить пакет Selenium по [ссылке](#). Обратите внимание, что на Вашем устройстве должен быть установлен браузер Google Chrome версии не ниже 85. После загрузки распакуйте файл и скопируйте путь до него, чтобы не потерять - например:

C:/Users/user/Desktop/chromedriver.exe

2.2 Установка пакетов

После этого откройте в Jupyter Notebook ipynb-файл с нашей программой и запустите установку дополнительных пакетов из первой ячейки: для этого необходимо нажать на ячейку с установкой пакетов (Рисунок 1) и запустить ее через Run.

```
In [17]: !pip install selenium
import time
from selenium import webdriver as wb
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.keys import Keys
import requests
import pandas as pd
from bs4 import BeautifulSoup
import warnings
warnings.filterwarnings('ignore')

Requirement already satisfied: selenium in c:\users\user\anaconda3\lib\site-packages (3.141.0)
Requirement already satisfied: urllib3 in c:\users\user\anaconda3\lib\site-packages (from selenium) (1.25.9)
```

Рис. 1: Ячейка с дополнительными пакетами

После того, как все пакеты загрузились (это можно понять по изменению цвета кружка в правом верхнем углу с черного на белый), переходите к следующему шагу.

2.3 Подготовка программы к работе

Перед тем, как запускать ячейку с программой, необходимо предварительно совершить два действия:

1. В первой же строчке программы изменить путь, по которому располагается драйвер Selenium, на тот, по которому Вы его распаковали (см. первый пункт). Строка выделена на рисунке ниже.

```
def declarator():
    br = wb.Chrome(r'C:\Users\user\Desktop\smstff\chromedriver.exe')
    br.get('https://google.com')
    x = input('Введите ФИО интересующего Вас чиновника: ') + ' ' + 'declarator.org/person -charts '
    search = br.find_element_by_name('q')
    search.send_keys(x)
    search.send_keys(Keys.RETURN)
```

Рис. 2: Строка с путем драйвера Selenium

2. (Если вы не собираетесь выгружать полученные данные в excel, то можете пропустить этот пункт.) В одной из последних строчек программы необходимо изменить название будущего датафрейма - по умолчанию он будет загружаться в ту же папку, где находится файл программы, поэтому необходимо, чтобы название файла не дублировалось с уже существующими там файлами такого же формата. Строчка так же выделена на рисунке ниже.

```
df2.style.set_properties(**{'text-align': 'left'})
f = input('Хотите ли вы скачать данные в excel-формате? Ответьте да или нет: ')
if f == 'да':
    df2.to_excel("df_dec4.xlsx")
    return df2
```

Рис. 3: Строчка с будущим названием файла excel

2.4 Работа программы

Для запуска программы повторите действия из пункта 2: нажмите на ячейку, после этого запустите её через Run. При запуске программы у вас откроется новое окно браузера Chrome. Рекомендуем развернуть его (перевести в полноэкранный режим), затем свернуть и забыть.

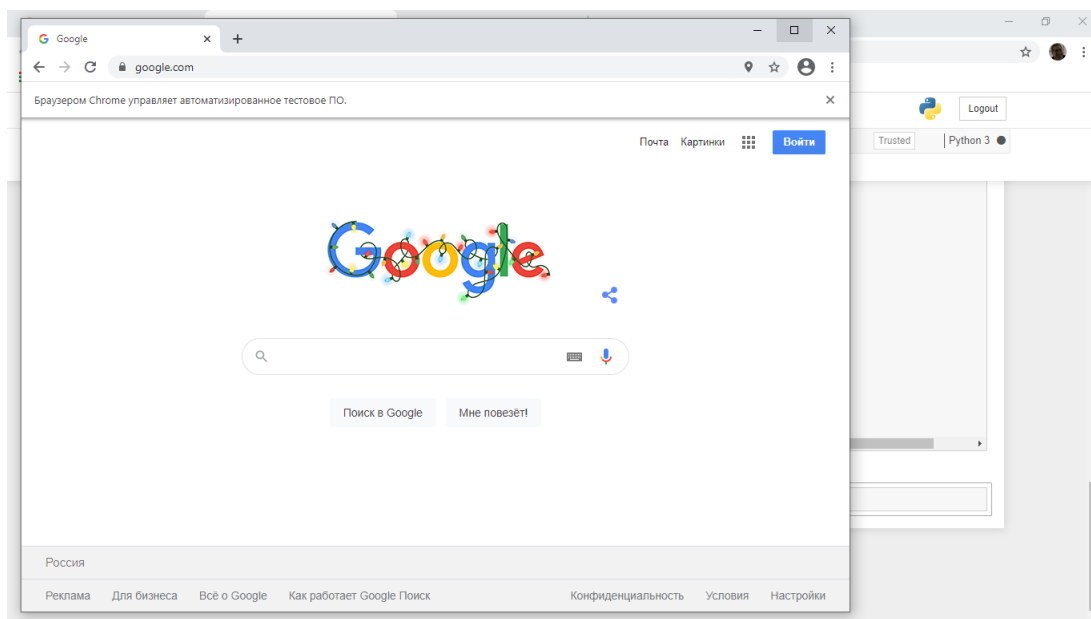


Рис. 4: Окно браузера, открытое программой

После того, как страница прогрузится, вам будет предложено ввести ФИО интересующего вас чиновника. Нажмите на появившееся поле и введите ФИО (можете ввести только фамилию и имя, но в таком случае есть риск, что программа предоставит неверные данные). После ввода перепроверьте ФИО на наличие опечаток и нажмите Enter.

Примерно через 20 секунд Вам будет предложено скачать данные в формате excel. Если вам необходимо выгрузить данные в excel формате, введите "да" (строго маленькими буквами), если Вам достаточно данных в блокноте - введите "нет". После ввода нажмите Enter.

```

else:
    return df2

declarator()

```

Введите ФИО интересующего Вас чиновника:

Рис. 5: Ввод данных о чиновнике

```

declarator()

```

Введите ФИО интересующего Вас чиновника: Глеб Никитин

Хотите ли вы скачать данные в excel-формате? Ответьте да или нет:

Рис. 6: Выгрузка данных в excel

После этого Вы увидите дата-фрейм со значениями дохода по всем имеющимся декларациям, площадью недвижимости, а также количеством транспорта. Последний столбец "Дополнительно" предоставляет более подробную информацию о доходах: доходы супруга или супруги, детей, марки машин и так далее. Чтобы подробнее с ней ознакомиться, откройте файл excel - он появится в той же папке, что и наша программа.

Out[5]:

	Декларация, статус, учреждение	Доход, руб.	Недвижимость, м2	Транспорт, шт.	Дополнительно
0	Антикоррупционная декларация 2019 Губернатор ...	4869247	2135	7	Доход 4 869 246,76 руб. Супруг(а): 3 241 168,7...
1	Антикоррупционная декларация 2018 Губернатор ...	4692512	2027	4	Доход 4 692 512,11 руб. Супруг(а): 5 885 506,7...
2	Антикоррупционная декларация 2017 Временно ис...	15737696	2027	4	Доход 15 737 696,32 руб. (в т.ч. доход, получе...
3	Антикоррупционная декларация 2016 Первый заме...	8323749	5690	4	Доход 8 323 748,80 руб. Супруг(а): 25 669 281,...
4	Антикоррупционная декларация 2016 Временно ис...	8323749	5833	4	Доход 8 323 748,80 руб. Супруг(а): 25 669 281,...
5	Антикоррупционная декларация 2015 Первый заме...	6789479	1981	3	Доход 6 789 479,06 руб. Супруг(а): 17 207 848,...
6	Антикоррупционная декларация 2014 Первый заме...	7250349	1877	3	Доход 7 250 349,02 руб. Супруг(а): 12 405 322,...
7	Антикоррупционная декларация 2013 Первый заме...	6336395	1946	3	Доход 6 336 394,55 руб. Супруг(а): 12 300 542,...

Рис. 7: Итоговый дата-фрейм

3 FAQ

3.1 Программа не работает, что не так?

Мы включили в код некоторые индикаторы, которые помогают понять, где именно возникла ошибка. Вместо итогового дата-фрейма Вы можете увидеть перед собой два типа сообщений:

1. *'Пожалуйста, перезапустите программу.'*

В данном случае программа не нашла ссылку на чиновника. Наиболее вероятная причина - изменение структуры кода страницы. Попробуйте перезапустить программу, переведя открывшееся окно в полноэкранный режим. Менее вероятно, но также возможна опечатка в ФИО чиновника: перепроверьте введенные

данные и так же перезапустите программу.

2. *'К сожалению, мы не смогли найти деклараций о доходах данного чиновника. Попробуйте найти декларации другого чиновника.'*

В данном случае структура кода точно не нарушена и точная причина - введенного Вами чиновника не существует в базе данных declarator.org. Перепроверьте введенные данные или введите ФИО другого чиновника.

3.2 Я хочу анализировать данные в таблице прямо в блокноте, как это сделать?

В данном случае после всех описанных шагов вам необходимо создать новую строку в блокноте и вписать туда код, показанный на рисунке ниже.

```
In [ ]: df = declarator()  
        print(df)
```

Рис. 8: Сохранение дата-фрейма в отдельную переменную

После этого просто повторите все шаги, которые были сделаны ранее. Отличие состоит в том, что теперь таблица будет сохранена в переменной df, с которой можно будет работать прямо в блокноте.