

**Московский государственный технический
университет им. Н.Э. Баумана**

**Факультет ИУ Кафедра
ИУ5**

**Курс «Основы информатики»
Отчет по домашнему заданию**

Выполнил студент группы ИУ5-33Б: Емельянов А.К.

Подпись и дата:

Проверил преподаватель каф.: Гапанюк Ю. Е.

Подпись и дата:

Москва, 2024 г

Цель: написать программу, которая будет сохранять файл с html кодом страницы плейлиста из сервиса Яндекс Музыка, для дальнейшего парсинга и внедрения данной функции в Телеграм бота.

Трудность с которой пришлось столкнуться: реклама, которая выскакивает при открытии страницы и встроенная защита сервиса от ботов. Для устранения этих помех был использован метод имитации живого пользователя с помощью задержки между действиями на страницу с помощью библиотеки `time`, и функции поиска и нажатия (`click()`) из библиотеки `selenium` для обхода всплывающей рекламы при открытии страницы.

КОД

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver import ActionChains
import time
def pars():
    url = "https://music.yandex.ru/users/alexemel2/playlists/3"

    options = webdriver.ChromeOptions()

    driver = webdriver.Chrome()

    try:
        driver.get(url=url)
        driver.maximize_window()
        time.sleep(3)
        cancel_button = driver.find_element(By.XPATH, "/html/body/div[1]/div[22]/div/span").click()

        time.sleep(3)
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        time.sleep(3)
        with open("/Users/act0n/Documents/proga/dz/pars_code_YANDEX_music.html", "w") as file:
            file.write(driver.page_source)

    time.sleep(50)
```

```
except Exception as ex:
    print(ex)
finally:
    driver.close()
    driver.quit()

pars()
```

Использование в боте

```
import telebot
from telebot.types import ReplyKeyboardMarkup, KeyboardButton
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver import ActionChains
import time
from lxml import html

bot = telebot.TeleBot('7715967994:AAFwur81tcHy2ap9eTh-_RVOA6mT3i7QZq8')
@bot.message_handler(commands=['start'])
def start(message):

    markup = ReplyKeyboardMarkup(resize_keyboard=True)
    button1 = KeyboardButton("регистрация")
    button2 = KeyboardButton("пришли html файл плейлиста")

    markup.add(button1, button2)

    bot.send_message(message.chat.id, "Выберите опцию:", reply_markup=markup)

name = ""
surname = ""
age = 0
@bot.message_handler(content_types=['text'])
def handle(message):

    if message.text == 'регистрация':
        bot.send_message(message.from_user.id, "Как тебя зовут?")
        bot.register_next_step_handler(message, get_name) #следующий шаг – функция get_name
    elif message.text=='пришли html файл плейлиста':
```

```

    bot.send_message(message.from_user.id, 'пришли ссылку на плейлист')
    bot.register_next_step_handler(message, pars)
else:
    bot.send_message(message.from_user.id, 'Напиши регистрация')

def get_name(message):
    global name
    name = message.text
    bot.send_message(message.from_user.id, 'Какая у тебя фамилия?')
    bot.register_next_step_handler(message, get_surname)

def get_surname(message):
    global surname
    surname = message.text
    bot.send_message(message.from_user.id, 'Сколько тебе лет?')
    bot.register_next_step_handler(message, get_age)

def get_age(message):
    global age
    while age == 0:
        try:
            age = int(message.text)
        except Exception:
            bot.send_message(message.from_user.id, 'Цифрами, пожалуйста')

    bot.send_message(message.from_user.id, 'Тебе '+str(age)+' лет, тебя зовут '+name+' '+surname+'?')

def pars(message):
    url = message.text
    options = webdriver.ChromeOptions()
    driver = webdriver.Chrome()
    print(url)

    try:
        driver.get(url=url)
        driver.maximize_window()
        time.sleep(3)
        cancel_button = driver.find_element(By.XPATH, "/html/body/div[1]/div[22]/div/span").click()

    time.sleep(3)

```

```

driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
time.sleep(3)
with open("/Users/acti0n/Documents/proga/dz/pars_code_YANDEX_music.html", "w") as file:
    file.write(driver.page_source)

time.sleep(1)
except Exception as ex:
    print(ex)
finally:
    driver.close()
    driver.quit()
    send_html(message)
    time.sleep(1)
    pars_html(message)

def send_html(message):
    try:
        with open("/Users/acti0n/Documents/proga/dz/pars_code_YANDEX_music.html", "rb") as file:
            bot.send_document(message.from_user.id, file, caption="Вот ваш HTML файл")
    except Exception as ex:
        bot.send_message(message.from_user.id, f"Произошла ошибка при отправке файла: {str(ex)}")

def pars_html(message):
    try:
        with open('/Users/acti0n/Documents/proga/dz/pars_code_YANDEX_music.html', 'r', encoding='utf-8') as file:
            content = file.read()
            bot.send_message(message.from_user.id, 'Плейлист найден, сейчас пришлю первые 10 песен')
    except Exception as ex:
        bot.send_message(message.from_user.id, f"Произошла ошибка при отправке файла: {str(ex)}")
    time.sleep(1)
    tree = html.fromstring(content)
    for i in range(1, 10):
        xpath_query = f"/html/body/div[1]/div[16]/div[2]/div/div/div[4]/div/div/div/div[1]/div[{i}]/div[2]/div[1]/div[1]/a"
        element = tree.xpath(xpath_query)

        # Проверка и извлечение текста
        if element and len(element) > 0:
            text = element[0].text_content() # Извлечение текста из элемента

```

```

else:
    print("Песня не найдена.")

xpath_query1 = f"/html/body/div[1]/div[16]/div[2]/div/div/div[4]/div/div/div/div[1]/div[{i}]/div[2]/div[1]/div[2]/span/a"
element1 = tree.xpath(xpath_query1)
if element1 and len(element1) > 0:
    text1 = element1[0].text_content() # Извлечение текста из элемента
else:
    print("Песня не найдена.")

bot.send_message(message.from_user.id, text+'by '+text1)
bot.polling(none_stop=True, interval=0)

```

результат

