Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной инженерии и компьютерной техники

**Отчёт**

**Лабораторная работа №2**

**Apache Spark.**

**Ф.И.О. студента** Герасимчук Михаил Юрьевич

**Группа** P4141

**Преподаватель** Русак А.В.

***г. Санкт-Петербург, 2022***

**Оглавление**

[Задание 3](#_Toc115820297)

[Цель: 3](#_Toc115820298)

[Основные этапы вычисления 3](#_Toc115820299)

[Сравнительный анализ методологий 3](#_Toc115820300)

[Iterative model 4](#_Toc115820301)

[Вывод 4](#_Toc115820302)

# Задание

1.Сбор данных с использованием API:

Осуществить сбор данных с любого сайта, предоставляющего API-интерфейс. Проанализировать состав данных, которые могут быть получены с помощью данного API-интерфейса.

2. Сбор данных без использования API:

Осуществить сбор данных с любого сайта, непредоставляющего API-интерфейс, с помощью инструментов парсинга сайтов. Проанализировать возможные варианты использования полученной информации.

# Сбор данных с использование API

Для получения данных при помощи API было решено использовать социальную сеть Вконтакте. Данная социальная сеть обладает открытым API, что даёт возможность любому пользователю получать и отправлять данные из социальной сети. Для использования VKAPI нам нужно было зарегистрировать приложение внутри сети, которое даст нам VK\_token. VK\_token, позволяет работать с VKAPI без авторизации.

Попробую написать запрос и получить данные о людях, чьи api нам известны.

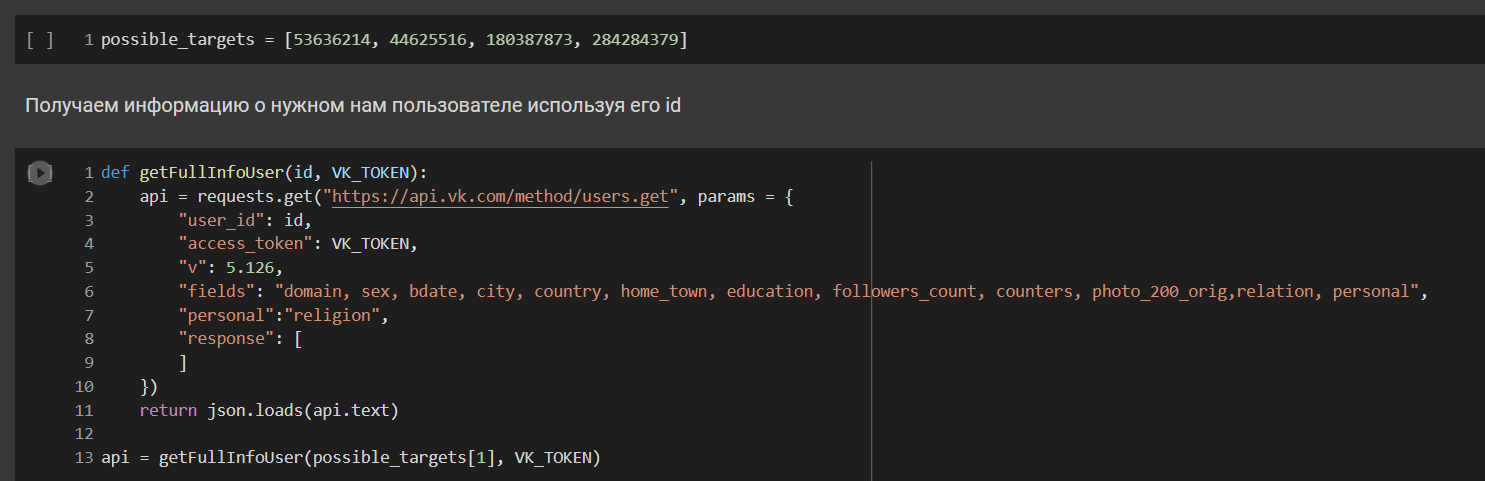


Рис. 1 Получение данных указанных на странице у нужных нам пользователей

После запроса нам возвращается файл в json формате, который в дальнейшем мы можем обрабатывать или анализировать.

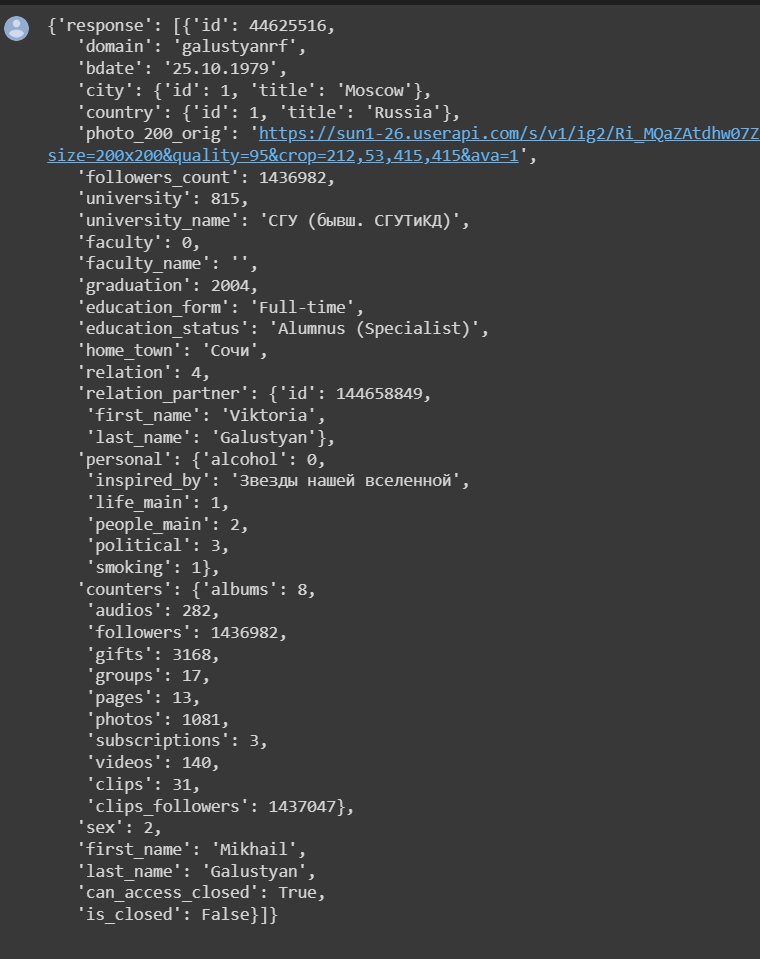


Рис. 2 Информация со страницы Михаила Галустяна

Также мы можем узнать id групп или друзей с которыми взаимодействует пользователь.

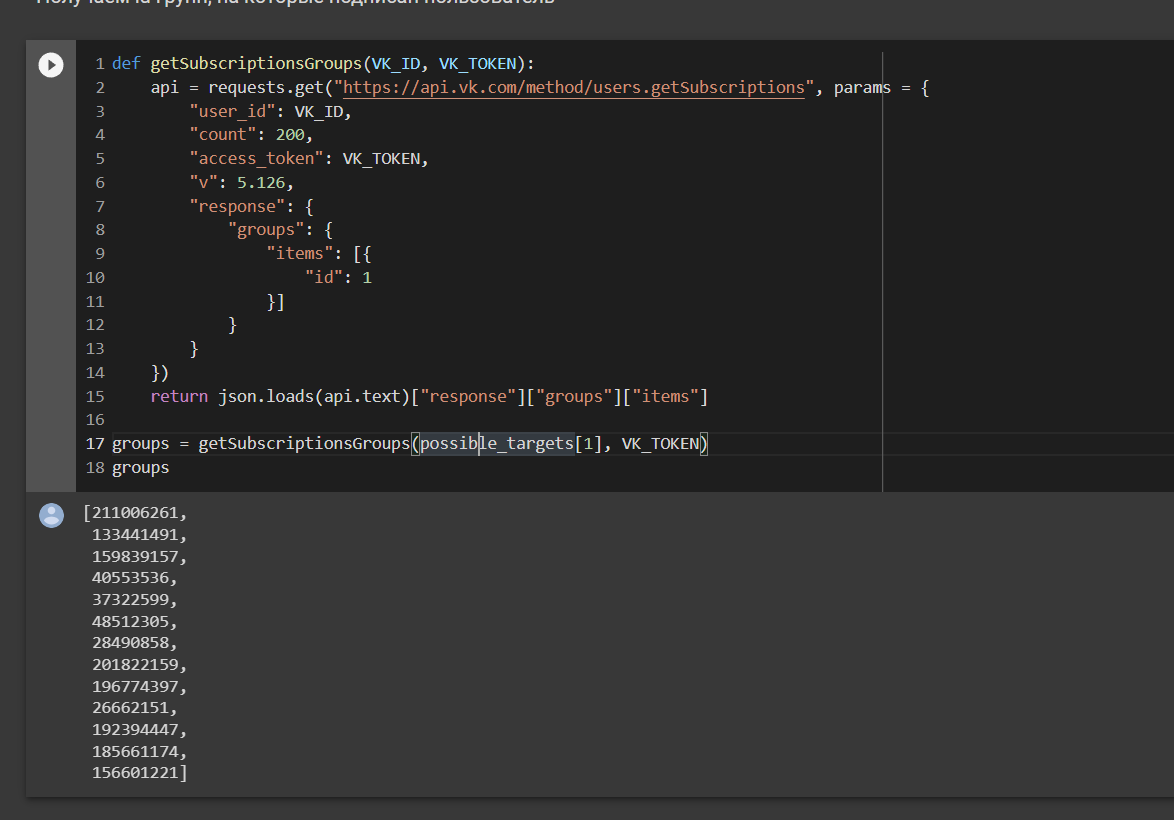


Рис. 3 Id групп на которые подписан Михаил Галустян

# Сбор данных без использования API

Для парсинга данных с сайта был выбран сайт Института международных отношений МИФИ. На сайте представлен список людей, которые ищут работы и мы попробуем получить их персональные данные. Всего на сайте представлены вакансии 347 человек. Нашу работу можно описать следующими пунктами:

1. Получение нужной веб страницы.

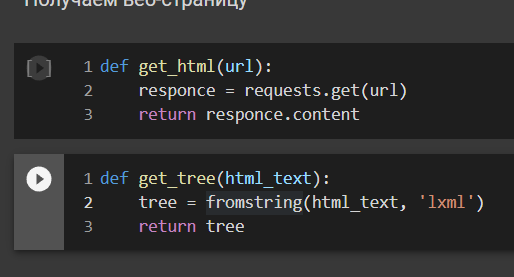


Рис. 1 Код для получения веб страницы указанной в запросе

1. Получение персональных данных

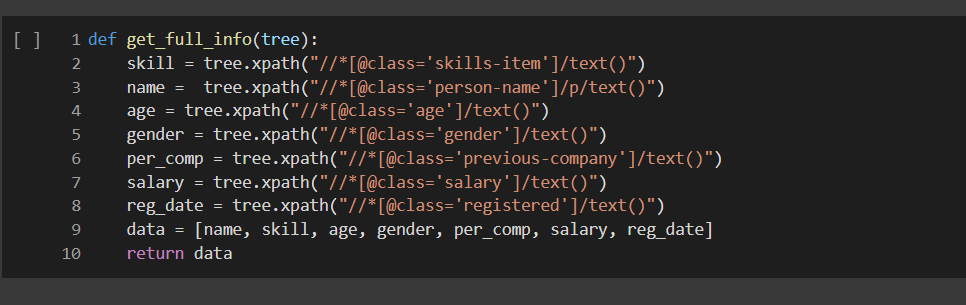


Рис. 2 Код для получения персональных данных указанных на сайте

1. Структурирование и сохранение данных

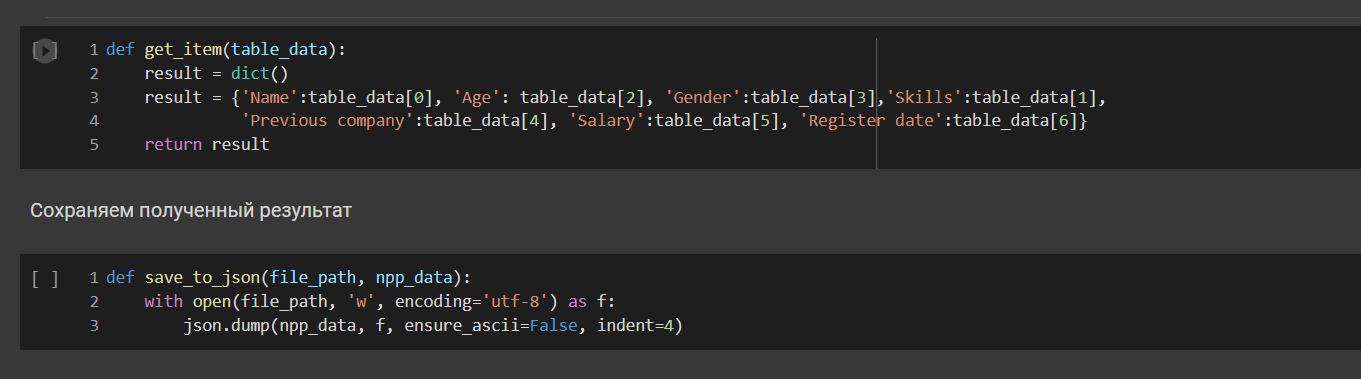


Рис. 3 Код для структурирования и сохранения полученных данных

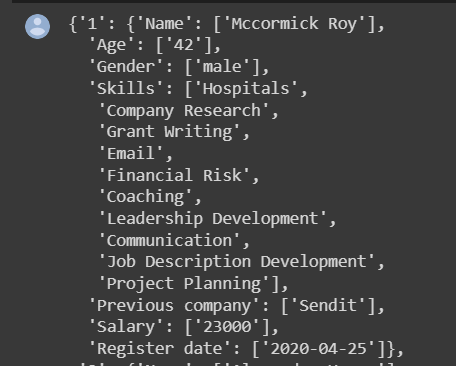


Рис. 4 Пример полученных данных

# Вывод

В процессе выполнения лабораторной работы мы научились извлекать данные из социальных сете и сайтов с использование и без использования API.