

**Министерство образования и науки Российской Федерации**  
**Федеральное государственное автономное образовательное учреждение высшего**  
**образования**  
**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ**  
**«МИФИ»**

---

**ИНСТИТУТ ФИНАНСОВЫХ ТЕХНОЛОГИЙ И ЭКОНОМИЧЕСКОЙ**  
**БЕЗОПАСНОСТИ**

**КАФЕДРА ФИНАНСОВОГО МОНИТОРИНГА**

**ОТЧЕТ**  
**ПО ЛАБОРАТОРНОЙ РАБОТЕ №3**  
**Методы извлечения больших данных**  
**ПО КУРСУ «РИАС»**

Выполнил \_\_\_\_\_ Антипенко Д.А. (С22-712).

Преподаватель \_\_\_\_\_ Прохоров И.В.

Оценка \_\_\_\_\_ Дата \_\_\_\_\_

Москва 2026

## **Введение**

Цель работы: освоить методы работы с большими данными, используя сервисы API ВКонтакте и API Яндекс.Диск.

Для этого предлагается изучить API-запросы, протестировать платформу для отправки API-запросов, создать цикл для извлечения больших данных, применить поиск и кластеризацию в больших данных.

Задание:

1. Реализовать следующие запросы к данным социальной сети ВКонтакте через API с помощью HTTP-запроса:
2. Кто онлайн в сообществе «Центр карьеры АО «Ил»?
3. Собрать все идентификаторы пользователей в указанном сообществе.
4. Вывести все последние пять записей нашего сообщества iRootHelp.
5. Вернуть всех друзей в сообществе, которые онлайн. Построить гистограмму полученных данных API ВКонтакте
6. Работа с Яндекс Диск.

# Задание 1. Технология извлечения данных из социальной сети ВКонтакте

## Методы API социальной сети ВКонтакте

Методы API социальной сети ВКонтакте – имеется 30 методов. В данной лабораторной работе мы рассмотрим некоторые из них. Полный список методов доступен по ссылке <https://vk.com/dev/methods>

Создадим приложение:

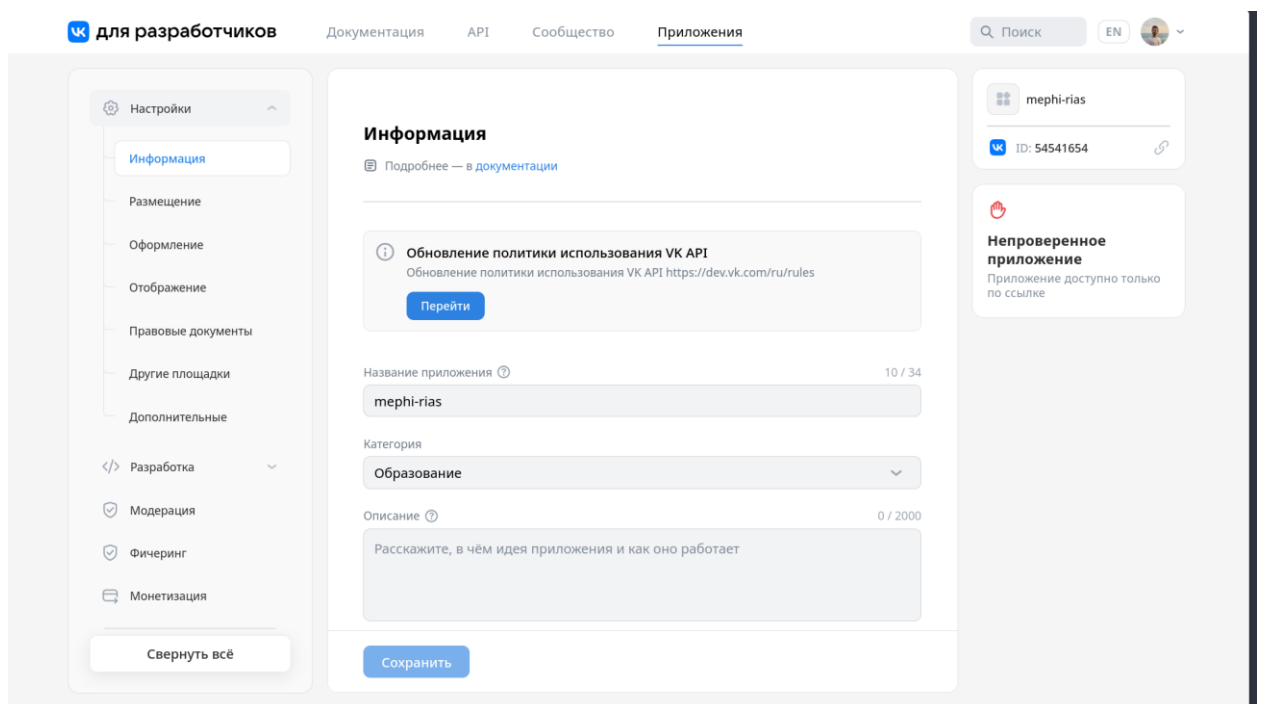


Рисунок 1

Затем привязываем устройство к аккаунту. После этого приложение будет создано, и откроется окно его настроек. Тут нам необходим идентификационный номер приложения и секретный ключ.

Далее с помощью запроса, мы получим code: **54541654**

[https://oauth.vk.com/authorize?client\\_id=54541654&redirect\\_uri=https://oauth.vk.com/blank.html&response\\_type=code](https://oauth.vk.com/authorize?client_id=54541654&redirect_uri=https://oauth.vk.com/blank.html&response_type=code)

<https://oauth.vk.com/blank.html#code=7099afa11f7112450c>

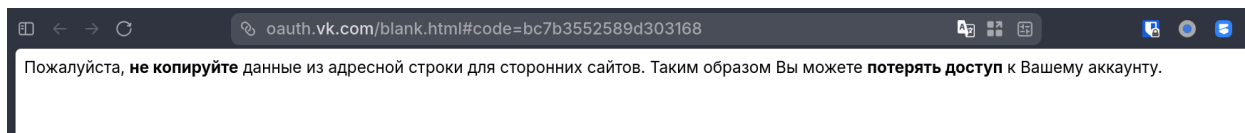


Рисунок 2

Теперь, когда у нас есть ID приложения и code, нужно выполнить следующий запрос в адресной строке браузера, чтобы поменять code на access\_token:

[https://oauth.vk.com/access\\_token?client\\_id=54541659&client\\_secret=sAeRZKn9TH1NZh0hqhYu&redirect\\_uri=https://oauth.vk.com/blank.html&code=52c3ed24847ba36fc1](https://oauth.vk.com/access_token?client_id=54541659&client_secret=sAeRZKn9TH1NZh0hqhYu&redirect_uri=https://oauth.vk.com/blank.html&code=52c3ed24847ba36fc1)

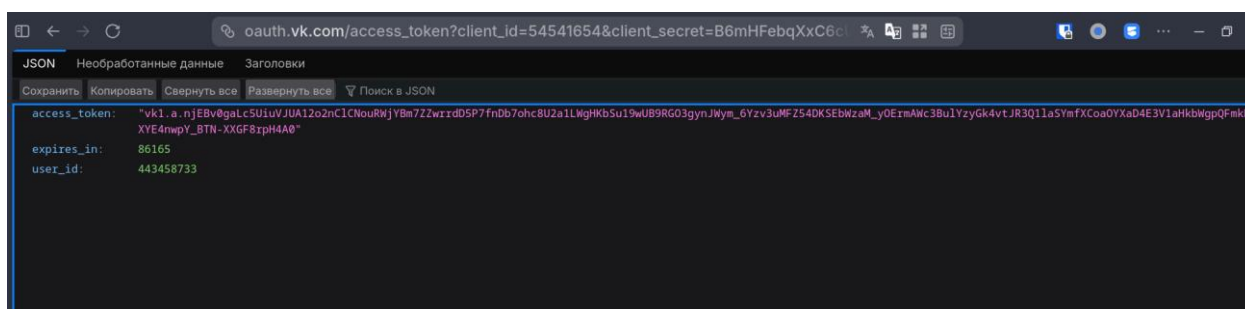


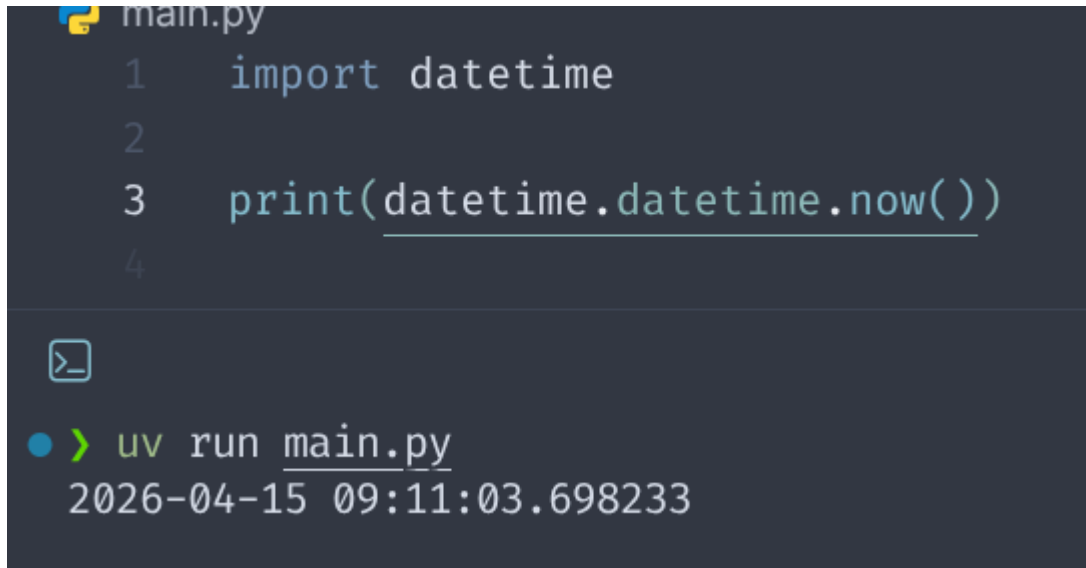
Рисунок 3

Содержимое страницы:

```
{
  "access_token": "vk1.a.njEBv0gaLc5UiuVJUA12o2nClCNouRWjYBm7ZZwrrdD5P7fnDb7ohc8U2a1LWgHKbSu19wUB9RGO3gynJWym_6Yzv3uMFZ54DKSEbWzaM_yOErmAWc3BulYzyGk4vtJR3Q1laSYmfXCoaOYXaD4E3V1aHkbWgpQFmkBDhLP-XYE4nwpY_BTN-XXGF8rpH4A0",
  "expires_in": 86165,
  "user_id": 443458733
}
```

Для работы с API был использован ЯП Python и среда разработки VSCode (на момент сдачи и публикации Лабораторной Работы, все приложения и ключи были удалены в целях приватности).

Была получена текущая дата:



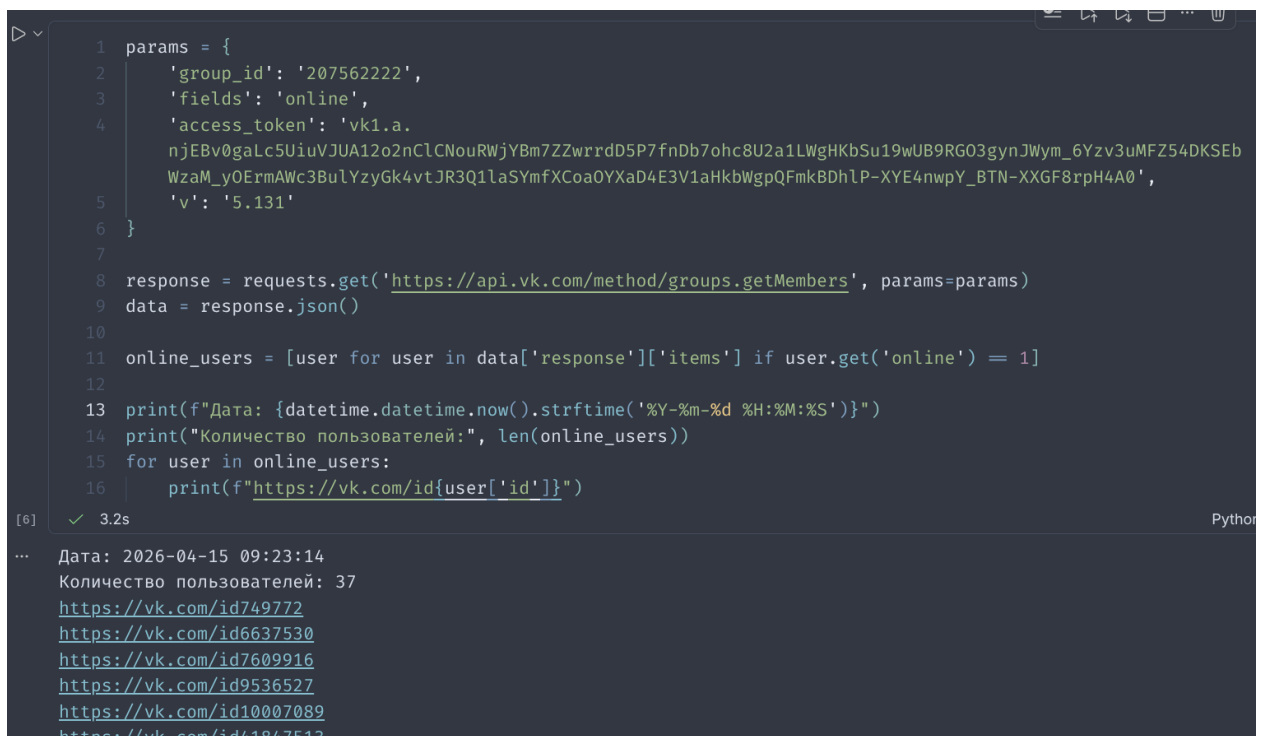
```
main.py
1 import datetime
2
3 print(datetime.datetime.now())
4

> uv run main.py
2026-04-15 09:11:03.698233
```

Рисунок 4

Будем рассматривать группу «Центр карьеры АО «Ил»  
group\_id=207562222.

- а) **Получить информацию о пользователях, находящихся онлайн в сообществе «Центр карьеры АО «Ил»:**



```
1 params = {
2     'group_id': '207562222',
3     'fields': 'online',
4     'access_token': 'vk1.a.
nJEbv0gaLc5UiuVJUA12o2nCLCNouRWjYBm7ZZwrrdD5P7fnDb7ohc8U2a1LWgHKbSu19wUB9RG03gynJWym_6Yzv3uMFZ54DKSEb
WzaM_y0ErmAWc3BulYzyGk4vtJR3Q1laSYmfXCoa0YXaD4E3V1aHkbWgpQFmkBDhLP-XYE4nwpY_BTN-XXGF8rpH4A0',
5     'v': '5.131'
6 }
7
8 response = requests.get('https://api.vk.com/method/groups.getMembers', params=params)
9 data = response.json()
10
11 online_users = [user for user in data['response']['items'] if user.get('online') == 1]
12
13 print(f"Дата: {datetime.datetime.now().strftime('%Y-%m-%d %H:%M:%S')}")
14 print("Количество пользователей:", len(online_users))
15 for user in online_users:
16     print(f"https://vk.com/id{user['id']}")
```

[6] ✓ 3.2s Python

... Дата: 2026-04-15 09:23:14  
Количество пользователей: 37  
<https://vk.com/id749772>  
<https://vk.com/id6637530>  
<https://vk.com/id7609916>  
<https://vk.com/id9536527>  
<https://vk.com/id10007089>  
<https://vk.com/id41847513>

Рисунок 5

С помощью программы на языке Python построим гистограмму распределения пользователей.

```
1 online_count = 0
2 offline_count = 0
3
4 for user in data['response']['items']:
5     if user.get('online') == 1:
6         online_count += 1
7     else:
8         offline_count += 1
9
10 print(f"Дата: {datetime.datetime.now().strftime('%Y-%m-%d %H:%M:%S')}")
11 labels = ['Онлайн', 'Оффлайн']
12 values = [online_count, offline_count]
13
14 plt.bar(labels, values, color=['blue', 'red'])
15 plt.title(f'Онлайн/оффлайн в группе {\'«Центр карьеры АО «Ил»\'})
16 plt.ylabel('Количество пользователей')
17 plt.show()
18
```

Рисунок 6

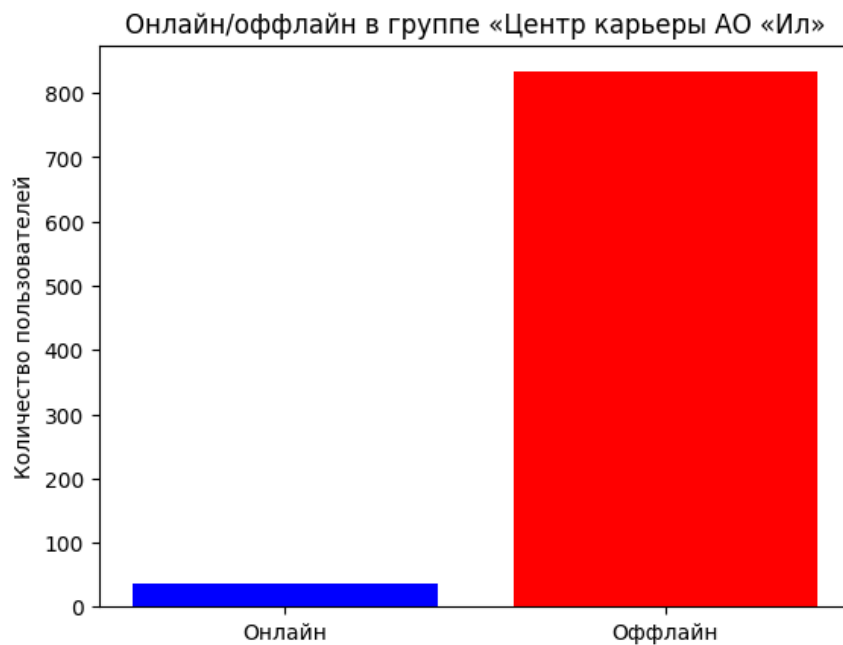
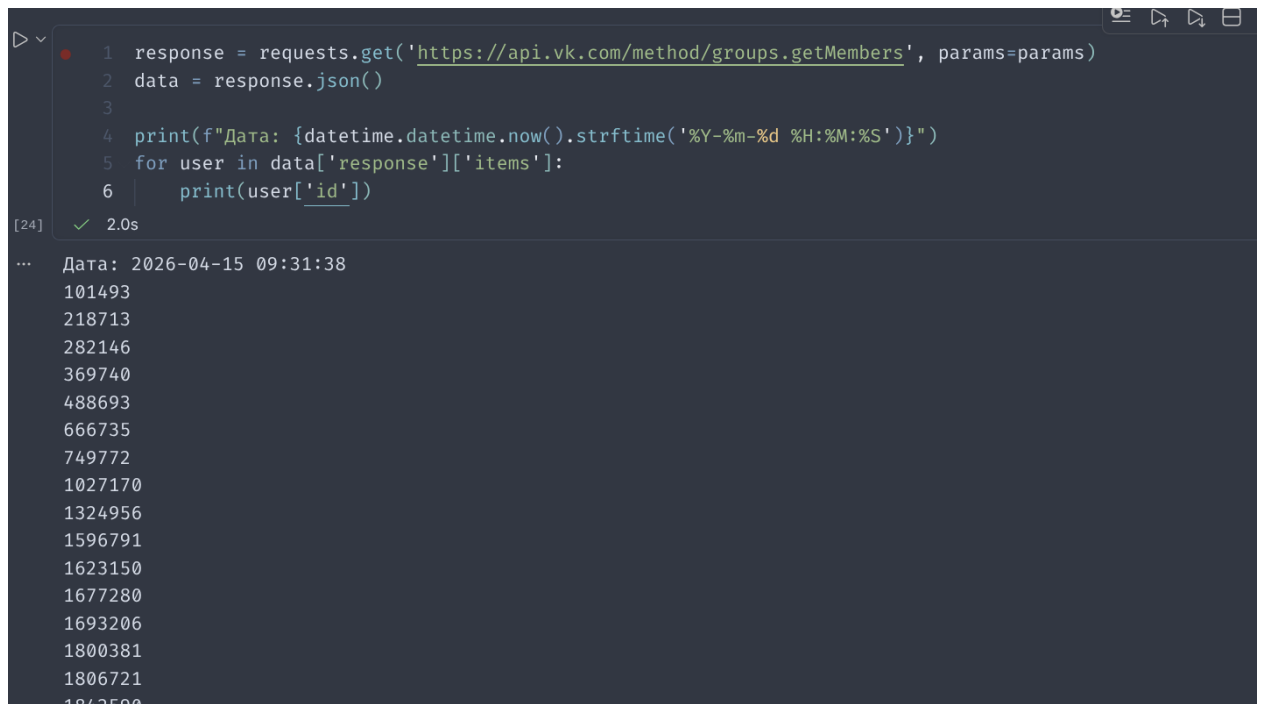


Рисунок 7

**б) Собрать все идентификаторы пользователей в указанном сообществе:**

Выполнение задания с помощью программы на языке Python:



```
1 response = requests.get('https://api.vk.com/method/groups.getMembers', params=params)
2 data = response.json()
3
4 print(f"Дата: {datetime.datetime.now().strftime('%Y-%m-%d %H:%M:%S')}")
5 for user in data['response']['items']:
6     print(user['id'])
```

[24] ✓ 2.0s

... Дата: 2026-04-15 09:31:38  
101493  
218713  
282146  
369740  
488693  
666735  
749772  
1027170  
1324956  
1596791  
1623150  
1677280  
1693206  
1800381  
1806721  
1842500

Рисунок 8

**с) Вывести последние пять записей сообщества:**

Выполнение задания с помощью программы на языке Python:

Код программы пункта 3 задания 1:

```
1 params1 = params.copy()
2 params1['owner_id'] = f'{-{207562222}}'
3 response = requests.get('https://api.vk.com/method/wall.get', params=params1)
4 data = response.json()
5
6 print(f"Дата: {datetime.datetime.now().strftime('%Y-%m-%d %H:%M:%S')}")
7 if 'response' in data:
8     posts = data['response']['items']
9     for post in posts:
10         print(f"Дата: {post['date']}, Текст: {post['text']}")
11 else:
12     print("Ошибка при получении данных:", data.get('error', {}).get('error_msg'))
13
```

✓ 0.3s Python

Дата: 2026-04-15 09:35:23  
Дата: 1773325922, Текст: Рассматриваете целевое обучение АО «Ил» для поступления? 🤔🤔

Что мы предлагаем:  
Форма обучения - очно, специалитет или бакалавриат.  
Обучение за счет бюджетных средств, с обязательством работы на предприятии после завершения обучения не менее 3-х  
Место работы - г. Москва или г. Жуковский

Преимущества поступления по целевому обучению АО «Ил»:  
💙 стипендия предприятия от 20 тыс. руб. при обучении на «хорошо» и «отлично»  
💙 прохождение всех видов практики на предприятии  
💙 трудоустройство возможно уже с 1-го курса

Рисунок 9

**d) Вернуть всех друзей в сообществе, находящихся онлайн:**

Выполнение задания с помощью программы на языке Python:

```
1 response_group_members = requests.get('https://api.vk.com/method/groups.getMembers', params=params)
2 data_group_members = response_group_members.json()
3
4 if 'response' not in data_group_members or 'items' not in data_group_members['response']:
5     print("Не удалось получить список участников сообщества.")
6     exit()
7
8 print(f"Количество людей в группе: {len(data_group_members['response']['items'])}")
9
10 url_friends = 'https://api.vk.com/method/friends.get'
11 response_friends = requests.get(url_friends, params=params)
12 data_friends = response_friends.json()
13
14 if 'response' not in data_friends or 'items' not in data_friends['response']:
15     print("Не удалось получить список друзей.")
16     exit()
17
18 print(f"Количество друзей: {len(data_friends['response']['items'])}")
19
20 friend_ids = {friend['id'] for friend in data_friends['response']['items']}
21
22
23 online_friends_in_group = [
24     user for user in data_group_members['response']['items']
25     if user.get('online') == 1 and user['id'] in friend_ids
26 ]
27
```

Рисунок 10

```
23 online_friends_in_group = [  
24     user for user in data_group_members['response']['items']  
25     if user.get('online') == 1 and user['id'] in friend_ids  
26 ]  
27  
28 if online_friends_in_group:  
29     for user in online_friends_in_group:  
30         print(f"https://vk.com/id{user\['id'\]}")  
31 else:  
32     print("Онлайн-друзей в сообществе не найдено.")  
33
```

[33] ✓ 3.3s

... Количество людей группе: 870  
Количество друзей: 598  
<https://vk.com/id250278958>

Рисунок 11

## Задание 2. Работа с облачным сервисом

### Яндекс.Диск

Технологии работы с данными, хранящимися в облачном хранилище из приложения Яндекс.Диск, позволяют управлять ресурсами Диска (файлами и папками) посредством пользовательского интерфейса и с помощью HTTP-запросов REST API. Предварительно нужно зарегистрировать аккаунт в Яндексе.

Чтобы зарегистрировать своё приложение, нужно перейти по ссылке <https://oauth.yandex.ru/>.

Права нужны на запись и чтение диска (или одной директории).

15 апр. 2026 г.

rias-tmp



#### Управление приложением



Почта для оповещений  
antip.ada@yandex.ru



Настройки



Удалить



#### Запрашиваемые права

Яндекс.Диск REST API



- Чтение всего Диска
- Доступ к информации о Диске

Далее необходимо получить отладочный token. С помощью запроса:  
[https://oauth.yandex.ru/authorize?response\\_type=token&client\\_id=1ad47da0eaa0473dbcd8c3e430535534](https://oauth.yandex.ru/authorize?response_type=token&client_id=1ad47da0eaa0473dbcd8c3e430535534)

Получен следующий токен:

y0\_\_xDGgsyeBhjLsUAg28fMixfZSDpWdmD8nTnQQGx7mbffUhEVLA

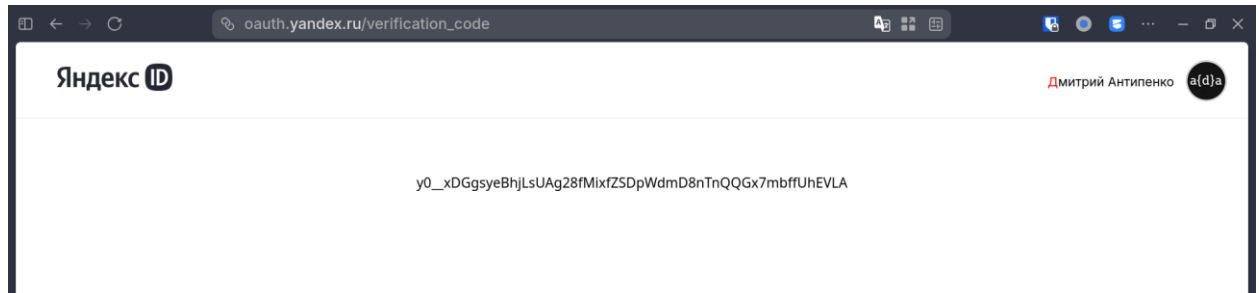


Рисунок 13

Задание выполнено на языке программирования Python:

## 1. Получить метainформацию о диске

Код программы:

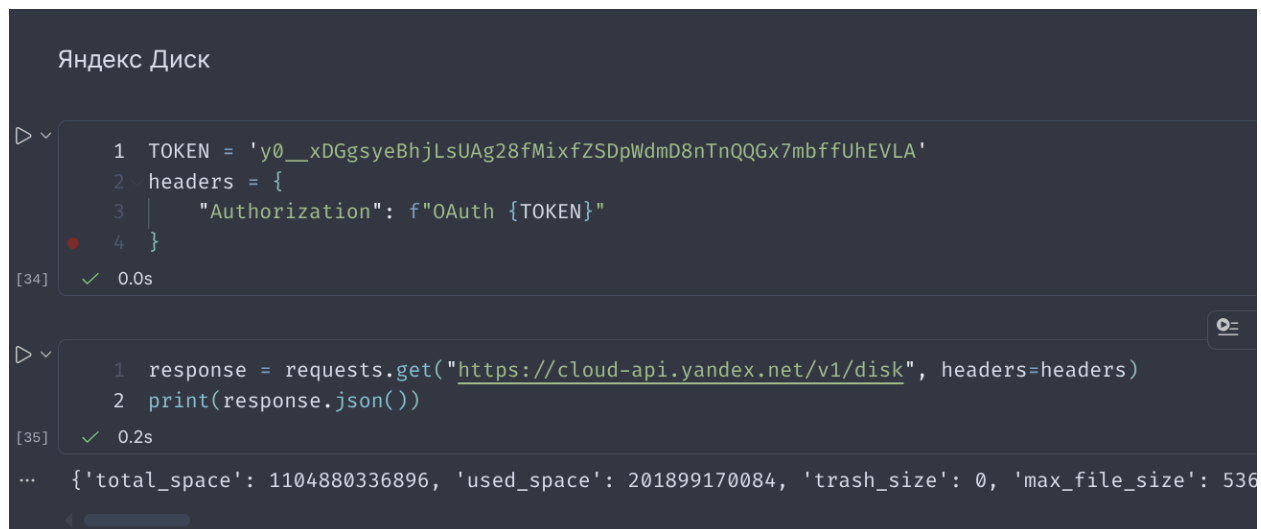


Рисунок 14

## 2. Получить метainформацию о файле или каталоге



Рисунок 15

### 3. Создать папку

```
▶ ▾  
● 1 params = {"path": "disk:/test_folder/rias"}  
  2 response = requests.put("https://cloud-api.yandex.net/v1/disk/resources",  
  3 print(response.status_code)  
[49] ✓ 3.9s  
... 201
```

Рисунок 16

Файлы

← test\_folder :



rias

Рисунок 17

### 4. Удалить файл или папку

```
▶ ▾  
1 params = {"path": "disk:/test_folder/rias"}  
2 response = requests.delete("https://cloud-api.yandex.net/v1/disk/resources", headers=headers,  
  3 print(response.status_code)  
[50] ✓ 0.7s  
... 204
```

Рисунок 18

← test\_folder :



Выберите файлы на компьютере  
или перетащите сюда

Загрузить файлы

Рисунок 19

## 5. Создать копию папки или файла

Заново создадим папку **rias** в папке **test\_folder** и скопируем её в корень диска.

```
1 params = {
2     "from": "disk:/test_folder/rias",
3     "path": "disk:/rias"
4 }
5 response = requests.post("https://cloud-api.yandex.net/v1/disk/resources/copy", headers=headers,
6                           params=params)
7 print(response.status_code)
```

[53] ✓ 1.1s

... 201

Рисунок 20

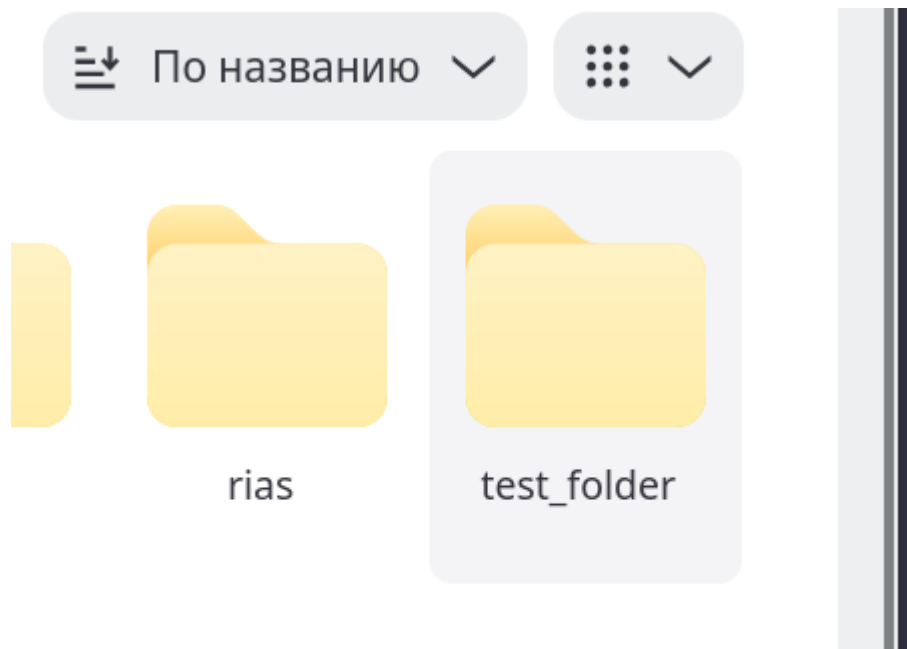


Рисунок 21

## 6. Получить ссылку на скачивание файла

```
1 params = {"path": "disk:/test_folder/rias"}
2 response = requests.get("https://cloud-api.yandex.net/v1/disk/resources/download", headers=headers,
3                           params=params)
4 print(response.json())
```

[54] ✓ 0.5s

Python

```
{'method': 'GET', 'href': 'https://downloader.disk.yandex.ru/zip/6af5ab8c80a9f5a11a1b763123c7e09cd8112cb3835abf0bc'}
```

Рисунок 22

## 7. Получить список упорядоченный по имени

```
1 params = {
2     "path": "disk:/test_folder/rias",
3     "sort": "name"
4 }
5 response = requests.get("https://cloud-api.yandex.net/v1/disk/resources", headers=headers, params=params)
6 data = response.json()
7
8 if '_embedded' in data and 'items' in data['_embedded']:
9     items = data['_embedded']['items']
10    for item in items:
11        print(item['name'])
12 else:
13    print("Файлы не найдены или папка пуста.")
```

[58] ✓ 0.4s Python

... penguin\_themed\_lock\_screen\_wallpaper\_with\_blue\_windows\_backdrop.jpg  
wavy\_lines\_v03\_5120x2880.png

Рисунок 23

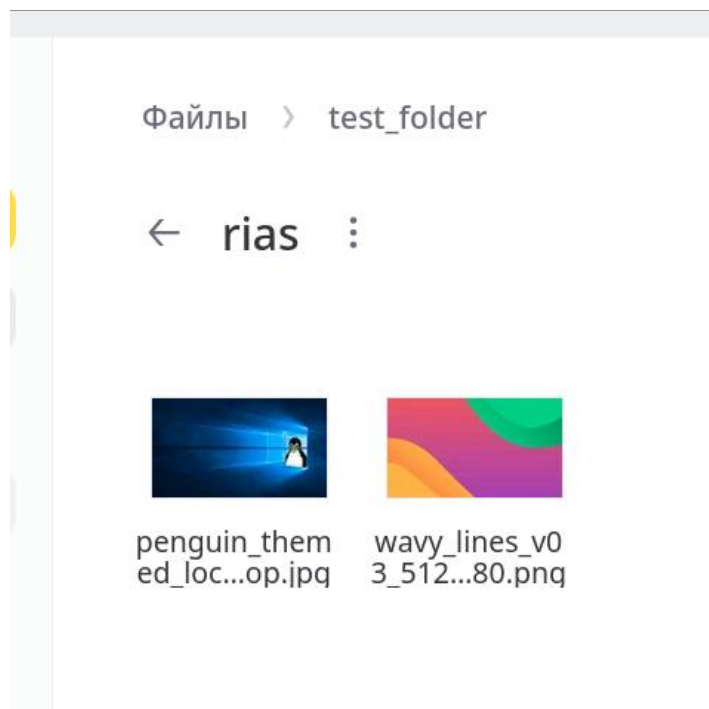


Рисунок 24

## 8. Получить список файлов упорядоченный по дате загрузки

```
1 response = requests.get("https://cloud-api.yandex.net/v1/disk/resources", headers=headers, params=params)
2 data = response.json()
3
4
5 if '_embedded' in data and 'items' in data['_embedded']:
6     items = data['_embedded']['items']
7     for item in items:
8         name = item.get('name')
9         created = item.get('created')
10        print(f"{name} - создан: {created}")
11 else:
12    print("Файлы не найдены или папка пуста.")
```

[57] ✓ 0.3s Python

... penguin\_themed\_lock\_screen\_wallpaper\_with\_blue\_windows\_backdrop.jpg - создан: 2026-04-15T07:03:08+00:00  
wavy\_lines\_v03\_5120x2880.png - создан: 2026-04-15T07:04:22+00:00

Рисунок 25

## 9. Переместить файл или папку

```
1 params = {
2     "from": "disk:/test_folder/rias/penguin_themed_lock_screen_wallpaper_with_blue_windows_backdrop.jpg",
3     "path": "disk:/test_folder/penguin_themed_lock_screen_wallpaper_with_blue_windows_backdrop.jpg",
4 }
5 response = requests.post("https://cloud-api.yandex.net/v1/disk/resources/move", headers=headers,
6 params=params)
7 print(response.status_code)
```

[59] ✓ 1.4s Python

... 201

Рисунок 26

Файлы

← test\_folder :

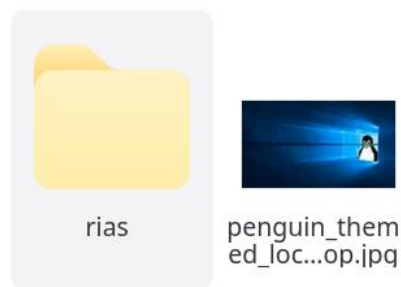


Рисунок 27

## 10. Получить список опубликованных ресурсов

```
1 response = requests.get("https://cloud-api.yandex.net/v1/disk/resources/public", headers=headers)
2 data = response.json()
3
4 if 'items' in data:
5     print(f"Список опубликованных ресурсов ({len(data['items'])}):")
6     for item in data['items']:
7         name = item.get('name', 'N/A')
8         public_url = item.get('public_url', 'N/A')
9         print(f"Имя: {name}, Ссылка: {public_url}")
10 else:
11     print("Опубликованные ресурсы не найдены.")
```

✓ 0.3s

Список опубликованных ресурсов (17):  
Имя: Ruby, Ссылка: [https://yadi.sk/d/\\_kCHdBAULZRAXQ](https://yadi.sk/d/_kCHdBAULZRAXQ)  
Имя: Лабы, Ссылка: <https://yadi.sk/d/4U5YW9JnqhHQDA>  
Имя: ОКНД(Андронов), Ссылка: <https://yadi.sk/d/gfy-aZipdTX4Uw>  
Имя: Проект-3, Ссылка: <https://yadi.sk/d/kFVeZA5ltSBwOA>  
Имя: Работы по МО, Ссылка: <https://yadi.sk/d/oML3-QgFo0vxsA>  
Имя: Философия, Ссылка: <https://yadi.sk/d/Eo4IrHcMwWQQzA>  
Имя: Форпост ИФТЭБ 22, Ссылка: [https://yadi.sk/d/424ii40\\_TboAmQ](https://yadi.sk/d/424ii40_TboAmQ)

Рисунок 28

## 11. Опубликовать ресурс

```
1 params = {"path": "disk:/test_folder/penguin_themed_lock_screen_wallpaper_with_blue_windows_backdrop.jpg"}
2 response = requests.put("https://cloud-api.yandex.net/v1/disk/resources/publish", headers=headers,
3 params=params)
4 print(response.json())
```

[62] ✓ 0.6s Python

... {'method': 'GET', 'href': 'https://cloud-api.yandex.net/v1/disk/resources?path=disk%3A%2Ftest\_folder%2Fpenguin\_the'}

Рисунок 29

## 12. Загрузить файл в диск по url

```
1 params = {
2     "url": "https://disk.yandex.ru/i/WbSw55_PxxFScg",
3     "path": "disk:/test_folder/rias/linux.jpg"
4 }
5 response = requests.post("https://cloud-api.yandex.net/v1/disk/resources/upload", headers=headers,
6 params=params)
7 print(response.json())
```

[63] ✓ 0.7s Python

... {'method': 'GET', 'href': 'https://cloud-api.yandex.net/v1/disk/operations/ba08a71edf20c5a86990329c1a475c21a1b5901'}

Рисунок 30

Файлы > test\_folder

← rias :



linux.jpg



wavy\_lines\_v0  
3\_512...80.png

Рисунок 31

### 13. Получить ссылку для загрузки файла

```
1 response = requests.get("https://cloud-api.yandex.net/v1/disk/resources/download", headers=headers,  
2 params=params)  
3  
4 download_url = response.json()["href"]  
5 print("Ссылка на скачивание:", download_url)  
6 file_response = requests.get(download_url, stream=True)  
7 file_response.raise_for_status()  
8  
9 local_file = Path("./penguin_themed_lock_screen_wallpaper_with_blue_windows_backdrop.jpg")  
10 with open(local_file, "wb") as out:  
11     for chunk in file_response.iter_content(chunk_size=8192):  
12         if chunk:  
13             out.write(chunk)  
14  
15 print(f"Файл скачан в: {local_file.resolve()}")
```

✓ 0.6s

Python

Ссылка на скачивание: <https://downloader.disk.yandex.ru/disk/a91c6b2b60dfc4b8c71920e03eeb1ebcb80b369bece72857fb4a9>  
Файл скачан в: [/home/ada/Documents/Files/files\\_mephi/Работы/РИАС/Lab3/penguin\\_themed\\_lock\\_screen\\_wallpaper\\_with\\_bl](/home/ada/Documents/Files/files_mephi/Работы/РИАС/Lab3/penguin_themed_lock_screen_wallpaper_with_bl)

Рисунок 32

## **Заключение**

В ходе работы я ознакомилась с основными методами извлечения данных из социальной сети ВКонтакте и Яндекс Диска с помощью API запросов. Для работы над этим были созданы приложения: ВК и Яндекс, а также получены соответствующие токены для авторизации.

На основе полученных знаний были осуществлены запросы по извлечению записей сообщества из социальной сети ВКонтакте, идентификаторов и списков участников сообщества. Для полученных с помощью API ВКонтакте данных о пользователях построена гистограмма распределения пользователей, находящихся онлайн и оффлайн.

В ходе выполнения заданий применены запросы, позволяющие вывести информацию о содержимом Яндекс Диска, а также запросы по управлению файлами и папками Яндекс Диска.