

Парсинг данных сообщества «ВКонтакте» с помощью Vк api

Вихляев Дмитрий Романович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Глаголев Владимир Александрович

Приамурский государственный университет имени Шолом-Алейхема

К.г.н., доцент кафедры информационных систем, математики и правовой информатики

Аннотация

Данная статья содержит описание и реализацию парсинга группы или сообщества в социальной сети «ВКонтакте». Программа реализована с помощью методов Api vk, и библиотеки jQuery. Результатом исследования станет готовая программа с считывающая данные о новостях сообщества.

Ключевые слова: vk api , парсинг, jQuery.

Sound speed control using the python programming language

Vikhlyayev Dmitry Romanovich

Sholom-Aleichem Priamursky State University

Student

Glagolev Vladimir Aleksandrovich

Sholom-Aleichem Priamursky State University

Candidate of Geographical Sciences, Associate Professor of the Department of Information Systems, Mathematics and Legal Informatics

Abstract

This article contains a description and implementation of the parsing of a group or community on the VKontakte social network. The program is implemented using the vk Api methods and the jQuery library. The result of the study will be a ready-made program that reads data about community news.

Keywords: vk api, parsing, jQuery.

1 Введение

1.1 Актуальность

Парсеры – это сервисы, которые помогают собрать и анализировать информацию из социальных сетей. Специальные программы анализируют код страницы с помощью различных алгоритмов. Фактически парсинг представляет процесс и инструмент поиска, по словам и выражениям, используя синтаксические правила языка применения. Результат

представляется, как правило, в виде структурированного файла или таблицы с возможностью дальнейшей обработки или визуализации полученной информации. Парсинг помогает тем, кому необходимо собирать и обрабатывать большие объёмы данных в интернете. Процесс парсинга не ограничивается одним сайтом, это могут быть личные сайты, сайты конкурентов и даже социальные сети. Парсеры собирают ID групп и личных страниц пользователей, и позволяют выгрузить эти данные в виде файла для последующей их загрузки. Парсинг очень часто применяется в рекламных целях. Например, для передачи пользователей соцсетей в кабинет в виде аудитории ретаргетинга.

1.2 Обзор исследований

Г.С.Сейдаметов, Ш.М.Усеинов, описали возможность загрузки фотографий через api vk на языке программирования php [1]. А.Г.Остапенко, Е.Р.Нежельский, М.Н.Степанов, Е.Ружицкий, А.Б.Заряев рассмотрели организацию мониторинга постов социальной сети ВКонтакте с помощью интерфейса vkarі [2]. В.А.Ильин, Д.А.Скоселев провели исследования парсинга веб-сайтов с использованием шаблонов [3]. Р.И.Круглик описал реализацию парсинга сайтов при помощи phpquery [4]. Д.К.Гончаров, Д.Д.Гончарова реализовали парсинг в соцсетях для политконсультанта [5]. Е.А.Байгушкина провела исследование алгоритма парсинга структур баз данных [6].

1.3 Цель исследования

Цель исследования – используя api vk и язык программирования javascript реализовать программу способную, извлекать данные, полученные от постов сообщества социальной сети «ВКонтакте» и заполнить таблицу соответствующими данными.

2 Материалы и методы

Для создания программы используется общедоступный интерфейс api vk и библиотека методов jQuery, языка программирования javascript.

3 Результаты и обсуждения

Чтобы пользоваться различными методами vk api, необходимо создать собственное приложение на основе аккаунта vk. Для этого нужно перейти на страницу <https://vk.com/dev>, начало работы в web. Далее зайти в каталог «мои приложения», нажать кнопку создать. Появится схема для заполнения данных. Название можно выбрать произвольное, его нигде не нужно будет указывать. Чтобы пользоваться api на личном домене, в качестве платформы берётся сайт, где указывается полный адрес сайта и базовый домен, который будет иметь доступ к методам api vk (рис. 1).

Название

Платформа: Встраиваемое приложение
 Standalone-приложение
 Сайт
 Скилл Маруси

Адрес сайта:

Базовый домен:

MyGrup
4 минуты назад
Звезда – массивное самосветящееся небесное тело, состоящее из газа или плазмы, в котором происходят, происходили или будут происходить термоядерные реакции. Ближайшей к Земле звездой является Солнце, другие звезды на ночном небе выглядят как точки различной яркости, сохраняющие своё взаимное расположение.

Yui – Rolling star 3:09

MyGrup
5 минут назад
Природа здесь верна стезе привычной,
Без ужаса берем удел обычный.
Но если вдруг, нежданная, вбегае
Беда в семью играющих Надежд.
Показать полностью...

Рис. 1. Создание приложения для получения данных сообщества

После создания приложения появится поле для изменения настроек приложения. Во вкладке настройки можно найти id приложения, он необходим для работы с vk api.

Для получения всех прав, необходимо получить access token. Для этого необходимо перейти в документацию, раздел знакомство с api «ВКонтакте», найти заголовок авторизация пользователя. В статье подробно указано как получить access token. Однако с новой версией api пришли изменения, которые не были указаны в документации. Пример действия новых стандартов будет разобран ниже.

После получения ключа доступа, можно загрузить данные сообществ «ВКонтакте». С помощью библиотеки «jQuery» и функции «ajax» вызывается запрос.

Первым параметром является url. В этот параметр указывается адрес в виде строки. Сначала идёт домен vk затем указывается метод. Для получения данных от сообществ в ВКонтате используется метод wall.get. После знака вопроса указываются параметры данного метода. Первый метод owner_id, является основным, присваивает идентификатор сообщества со знаком «-». Остальные методы являются не обязательными, но их нужно добавлять с помощью знака «&». В новых версиях api обязательно нужно указать версию данного api. На данный момент актуальными остаются версии 5.8 и выше. А также access token. Их также можно добавить с помощью знака «&».

Во второй параметр указывается метод, которым будет отправляться запрос.

Третий параметр «dataType», в него нужно указать «jsonp», для того чтобы данный запрос с клиента успешно был выполнен.

Дальше, если запрос к api vk будет успешен, будет вызвана функция, которая принимает в качестве параметров данные в виде массива, после чего вытаскивает и обрабатывает необходимые данные для дальнейшей работы с ними (рис.2,4).

```
function getUrl(method, params)
{
    if(!method) throw new Error('неуказан метод');
    params = params || { };
    params ['access_token'] =
        'e5097a87b6986cad616948d6a2fa5fa77e2149311cde2d25dc83837e9695fc7874e9ef7c8e3dc6226acc1';
    params ['v'] = '5.131';

    return 'https://api.vk.com/method/'+method+'?' + $.param(params);
}
function sendRequest(method, params, func)
{
    $.ajax({
        url:getUrl(method, params),
        type: 'GET',
        dataType: 'jsonp',
        success: func
    });
}
function loadNuws()
{
    sendRequest('wall.get',{owner_id: -207754709}, function (data){
        show_nuws(data.response.items);
    });
}
```

Рис. 2. Получения данных сообщества «ВКонтакте»

Весь список данных можно вывести в консоль браузера в виде объекта, внутри которого находится ассоциативный массив. Самый главный массив «response», в нём хранится количество элементов «count» и сам список новостей «items». В последнем, также хранится массив данных о конкретном посте. В данной статье используются элементы:

«attachments» данный массив хранит в себе прикрепленные к посту файлы, такие как фото, видео, музыка, документ и тд.

«date» хранит закодированную дату и время, в которую был опубликован пост.

«text» хранит текст, оставленный в данном посте (рис.3).

```

Object
  response:
    count: 13
    items: Array(13)
      0:
        attachments: (2) [{"..."}, {"..."}]
        can_delete: 1
        can_pin: 1
        carousel_offset: 0
        comments: {count: 0}
        date: 1633920764
        donut: {is_donut: false}
        from_id: -207754709
        hash: "F1GaDW65aphr658cXG_mOZ_--71L"
        id: 16
        is_favorite: false
        likes: {can_like: 0, count: 0, user_likes: 0}
        marked_as_ads: 0
        owner_id: -207754709
        post_type: "post"
        reposts: {count: 0, wall_count: 0, mail_count: 0}
        short_text_rate: 0.8
        text: "Звездá — массивное самосветящееся небесное тело, состоящее из газа или плазмы, в котор
        views: {count: 3}
        [[Prototype]]: Object
      1: {id: 15, from_id: -207754709, owner_id: -207754709, date: 1633920714, marked_as_ads: 0, ...}
      2: {id: 14, from_id: -207754709, owner_id: -207754709, date: 1633920371, marked_as_ads: 0, ...}
      3: {id: 13, from_id: -207754709, owner_id: -207754709, date: 1633920149, marked_as_ads: 0, ...}
      4: {id: 12, from_id: -207754709, owner_id: -207754709, date: 1633920072, marked_as_ads: 0, ...}
      5: {id: 10, from_id: -207754709, owner_id: -207754709, date: 1633848251, marked_as_ads: 0, ...}
    
```

Рис. 3. Вывод данных в консоль

Все имеющиеся данные можно отобразить в таблице. Каждой новости выделяется одна строка, а информация разбивается по столбцам (рис.4).

Date	Text	Photo	Video	Audio	Doc
Mon Oct 11 2021 12:52:44 GMT+1000 (Владивосток, стандартное время)	Звездá — массивное самосветящееся небесное тело, состоящее из газа или плазмы, в котором происходят термоядерные реакции. Ближайшей к Земле звездой является Солнце, другие звезды на ночном небе выглядят как точки различной яркости, сохраняющие свое взаимное расположение.	https://sun9-41.userapi.com/impz/ewH6evrep5UjZ523nD... STXU8aGh5vwhk1v-Q-3hcORZl0NpM.jpg?size=510x338&quality=96&crop=43,0,310,338&sign=01934336bd04ea6685c2e2527e79bd&cc_unix_tag=UJL8oGw_7iLZz63kgevx-DN8F4E9ZVj4-p9aAU&type=album			https://vk.com/mp3/audio_api_unavailable.mp3
Mon Oct 11 2021 12:51:54 GMT+1000 (Владивосток, стандартное время)	Природа здесь весьма своеобразна. Без ужаса берем удел обычный. Но если вдруг, нежданная, побегает Бега в сторону играющих Наюжи. Но если вдруг, внезапно слетает С веселых, ей лишь миг знакомых вежд И Счастие младое умиряет. Еле не свес и призрачные снажки... Тогда наш дуг облетает трепетом. И силой в грудь врывается роптанье.	https://sun9-52.userapi.com/impz/VTQzR8hgw7a_PvovT... sGFCq8pPAhIw_GUxU38g-oDDIM6MHQGA.jpg?size=510x382&quality=96&sign=07cece4c151a3a7f47c010710191634&cc_unix_tag=N4Xv0e...wArCYWmqKIBiA1a3eJqC5vcUPX7ou7suO&type=album			
Mon Oct 11 2021 12:46:11 GMT+1000 (Владивосток, стандартное время)	Окраина природы — значимый комплекс мер по сохранению, рациональному использованию и восстановлению природных ресурсов и естественной окружающей среды, в том числе выделено разнообразия флоры и фауны, богатства надр, чистоты вод, лесов и атмосферы Земли.	https://sun9-88.userapi.com/impz/oDj5p_OH_p9I2qC... r64w2kVx4r78qGVW4_Vjg4q644d8C.jpg?size=284x177&quality=96&sign=93686920464b779a65eeae5f8aF7e&cc_unix_tag=Mn62p6Ak...lJy6vkg4p2qXmVx0vTnd0Wx6Ajp4FJL&type=album			https://vk.com/doc386139966_6138636247.html?w=7... hasi=7e6e7...aa83bc5daafbd&dl=GM4DMJTFHE4T7INQ_1634033513-7378beb2bc8be542&api=1&no_preview=1
Mon Oct 11 2021 12:42:29 GMT+1000 (Владивосток, стандартное время)	Сновидение — субъективное восприятие образов, возникающих в сознании спящего человека. Сновидения происходят в основном в фазе быстрого сна, когда активность мозга высока и напоминает бодрствование.		https://vk.com/video-149562500_456239018 https://vk.com/video-149562500_456239020		
Mon Oct 11 2021 12:41:12 GMT+1000 (Владивосток, стандартное время)	Звездá — массивное самосветящееся небесное тело, состоящее из газа или плазмы, в котором происходят термоядерные реакции. Ближайшей к Земле звездой является Солнце, другие звезды на ночном небе выглядят как точки различной яркости, сохраняющие свое взаимное расположение.	https://sun9-41.userapi.com/impz/ewH6evrep5UjZ523nD... STXU8aGh5vwhk1v-Q-3hcORZl0NpM.jpg?size=510x338&quality=96&crop=43,0,310,338&sign=01934336bd04ea6685c2e2527e79bd&cc_unix_tag=UJL8oGw_7iLZz63kgevx-DN8F4E9ZVj4-p9aAU&type=album			https://vk.com/mp3/audio_api_unavailable.mp3
Sun Oct 10 2021 16:44:11 GMT+1000 (Владивосток, стандартное время)			https://vk.com/video-54341096_167372724 https://vk.com/video199229897_456239186		https://vk.com/doc386139966_5068458587.html?w=7... hasi=7e6e7...aa83bc5daafbd&dl=GM4DMJTFHE4T7INQ_1634033513-7378beb2bc8be542&api=1&no_preview=1

Рис. 5. Таблица с данными о новостях

Таким образом, была написана программа для парсинга данных из новостей сообщества в социальной сети «ВКонтакте».

Библиографический список

1. Сейдаметов Г.С., Усеинов Ш.М. Загрузка фотографий через api vk на

- языке программирования php //Информационно-компьютерные технологии в экономике, образовании и социальной сфере. 2018. № 2 (20).
2. Остапенко А.Г., Нежелский Е.Р., Степанов М.Н., Ружицкий Е., Заряев А.В. Организация мониторинга постов социальной сети ВКонтакте с помощью интерфейса vкарі // Информация и безопасность. 2018. Т. 21. № 3. С. 408-415.
 3. Ильин В.А., Скоселев Д.А. Парсинг веб-сайтов с использованием шаблонов //Мир современной науки. 2018. № 2 (48). С. 8-12.
 4. Круглик Р.И. Парсинг сайтов при помощи phpquery // Постулат. 2020. № 1 (51). С. 62.
 5. Гончаров Д.К., Гончарова Д.Д. Парсинг в соцсетях для политконсультанта // Modern Science. 2020. № 5-2. С. 270-274.
 6. Байгушкина Е.А. Исследование алгоритма парсинга структур баз данных // Энигма. 2021. № 29-2. С. 132-136.
 7. Маринчук А. С. Выгрузка фотографий с личной страницы в социальной сети «ВКонтакте» с помощью vk арі //Постулат. 2021. № 1 (63).
 8. Галимов Р.Р., Стефанова И.А. Способы получения данных из социальных сетей для поиска потенциальных абитуриентов // Инновации в науке. 2016. № 57-1. С. 37-41.
 9. Крайников А.Н., Банников А.А. Управление "gріо" с помощью "vk-арі" // Международный журнал прикладных наук и технологий Integral. 2020. № 4. С. 34.