

**МИКРОСЕРВИС ДЛЯ ЗАГРУЗКИ ДАННЫХ
ИЗ СОЦИАЛЬНОЙ СЕТИ TWITTER ДЛЯ АНАЛИЗА ЭМОЦИЙ ПОЛЬЗОВАТЕЛЕЙ**

М.В. Мытницкая, Е.О. Шубкин.

Научный руководитель: доцент, к.т.н. Е.Е. Лунева

Национальный исследовательский Томский политехнический университет,

Россия, г. Томск, пр. Ленина, 30, 634050

E-mail: mashka7737@gmail.com

**MICROSERVICE FOR DOWNLOADING DATA
FROM SOCIAL NETWORK TWITTER FOR ANALYSIS OF USERS EMOTIONS**

M.V. Mytnitskaya, E.O. Shubkin

Scientific Supervisor: Associate Prof., Ph.D. E.E. Luneva

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: mashka7737@gmail.com

***Abstract.** Micro blogs on the Twitter platform have become an important tool for exchanging views and different points of view. The analysis of these opinions is a powerful tool that is applicable in many areas of activity. In particular, automated analysis of messages can provide hidden information about the state of the stock market. Studies of economists from the Munich Technical University showed that the moods that are present in messages from Twitter similarly develop in the stock market and moreover they sometimes outpace them for one day.*

Введение. Микро блоги на платформе Twitter стали важным инструментом для обмена мнениями и различными точками зрения. Анализ этих мнений является мощным инструментом, применимым во многих сферах деятельности. В частности, автоматизированный анализ сообщений может дать скрытую информацию о состоянии рынка акций. Исследования экономистов из Мюнхенского Технического Университета показали, что настроения, присутствующие в сообщениях из Twitter аналогично развиваются и на фондовом рынке и более того порой они опережают их на один день.

Цель работы. Разработка микросервиса для загрузки данных из социальной сети для последующего анализа предложений на основе разработанной методики.

Задачи.

1. Изучить API Twitter для загрузки массива данных, поиск сообщений пользователей по ключевым словам.
2. Реализация тестового сбора данных из социальной сети.
3. Изучить методики работы с использованием технологии MVC
4. Реализация микросервиса с собственным API для выгрузки данных по заданным пользователем параметрам.
5. Проанализировать способы автоматизированного анализа текста, рассмотреть грамматику составляющих, зависимости, дерева разбора.
6. Тестирование и развертывание микросервиса.

Материалы и методы исследования. REST API - архитектурный стиль взаимодействия компонентов распределённого приложения в сети.

В сети Интернет, вызов удалённой процедуры может представлять собой обычный HTTP-запрос (обычно «GET» или «POST»; такой запрос называют «REST-запрос»), а необходимые данные передаются в качестве параметров запроса. Отсюда следует главное правило REST архитектуры - в период между запросами клиента никакая информация о состоянии клиента на сервере не хранится. Все запросы от клиента должны быть составлены так чтобы сервер получил всю необходимую информацию для выполнения запроса [1].

Основные преимущества микросервисной архитектуры:

- Разработка и поддержка микросервисов небольшого размера во многом проще чем разработка и поддержка микросервисов больших размеров. Чем короче код, тем легче восстановить его в памяти.
- При использовании REST API становятся доступны все присущие ему преимущества такие как простота интерфейса, надёжность и легкость внесения изменений [2].
- Поскольку каждый микросервис представляет собой отдельный проект, это позволяет строить иерархии из сервисов. То есть, некие сервисы будут использованы только парой сервисов, предоставляющих внешний API для других сервисов. Масштабированный подход позволяет одновременно работать над одной системой сразу нескольким программистам. За счет этого исчезают проблемы, связанные с разрешением конфликтов или поиском сложных логических ошибок.

Результаты. Создана программа, которая выполняет запрос, с использованием Twitter API в социальную сеть.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ArrayOfTweets xmlns:i="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://schemas.datacontract.org/2004/07/...">
  <Tweets>
    <CreatedBy>dc</CreatedBy>
    <CreatedByID>1323</CreatedByID>
    <CreationDate>2016-12-23T23:52:03.2</CreationDate>
    <FullText>Hello</FullText>
    <Id>1</Id>
    <IsRetweet>false</IsRetweet>
    <RetweetCount>4</RetweetCount>
  </Tweets>
  <Tweets>
    <CreatedBy>Jondalf Thrasher</CreatedBy>
    <CreatedByID>800380580042944512</CreatedByID>
    <CreationDate>2016-12-23T23:52:00.397</CreationDate>
    <FullText>
      RT @SupermanTweets: Preparing for President Trump... https://t.co/eHGBzNI7L
    </FullText>
    <Id>2</Id>
    <IsRetweet>true</IsRetweet>
    <RetweetCount>20</RetweetCount>
  </Tweets>
  <Tweets>
    <CreatedBy>Eugenio Bertolaccini</CreatedBy>
    <CreatedByID>276544322</CreatedByID>
  </Tweets>
</ArrayOfTweets>
```

Рис. 1. Ответ микросервиса в виде сериализованных данных XML

Для обеспечения авторизации приложения для взаимодействия с API Twitter создан класс авторизации пользователя с использованием pin-кода. Из полученных данных исключаются ненужные атрибуты (т.е. все кроме текста твита, ID сообщения, имени автора, ID автора, оригинального поста и количественных характеристик - количество лайков, репостов). Используется собственный экземпляр класса, который агрегирует полученную выборку. При запросе пользователя к микросервису,

упорядоченные данные возвращаются в формате json или xml. При тестировании приложения возвращаемый формат зависит от браузера, из которого поступает запрос (рис. 1).

При разработке использовалась такая архитектура программы, которая позволила подключить разрабатываемые классы к новому микросервису без значительного изменения кода.

Микросервис предоставляет собственный API пользователю, выполняющий набор функций. Пользователь на вход микросервису отправляет ключевое слово или хэштег (или несколько ключевых слов).

Для тестирования микросервиса разработана простой графический интерфейс для обеспечения запросов к сервису (рис. 2).

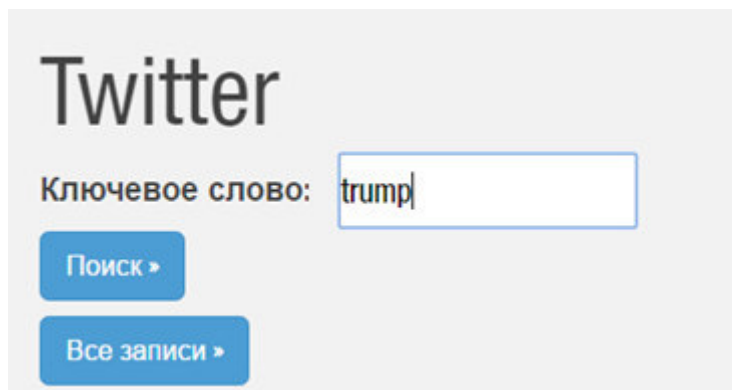


Рис. 2. Графический интерфейс для тестирования микросервиса

Итоговая база данных содержит в себе информацию о отправителе, тексте сообщения и других дополнительных параметрах.

Созданный микросервис позволяет совершить 15 запросов в течении 15 минут не авторизованному пользователю и 180 запросов за 15 минут авторизованному пользователю.

Один запрос может содержать не более 100 записей (стоит отметить что Twitter предоставляет данные о пользователях и твитах только за последние 6-9 дней).

Планы. Произвести изменение типа авторизации пользователя на URL redirect authentication. Осуществить переход на систему управления базами данных MongoDB. Далее в планах создание нейросети для анализа эмоций пользователей.

Выводы. Исследовательская часть работы по анализу данных связана с изучением теории и разработкой методики анализа текста. Основной результат – готовый микросервис для загрузки данных для последующего анализа предложений на основе разработанной методики.

СПИСОК ЛИТЕРАТУРЫ

1. Лунёва Е.Е. Способ оценки эмоций пользователей с использованием нечеткой логики на примере социальной сети Twitter [Электронный ресурс]. - режим доступа: <http://elibrary.ru/item.asp?id=23598969>. – 21.10.16
2. Hradil J., Sklenak V. (2017) Practical Implementation of 10 Rules for Writing REST APIs//Journal of Systems Integration (1804-2724), 8(1), 45-54.