

УДК 004.42

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ АНАЛИЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ С ПРИМЕНЕНИЕМ ЭЛЕМЕНТОВ ГЕОСТАТИСТИКИ

Инсапов А.Р, студент гр. ПРО-ИТ-201М, 2 курса
Научный руководитель: Воробьева Г.Р., д.т.н., доцент
Уфимский университет науки и технологии
г. Уфа

Аннотация.

В статье представлено стек используемых технологий и методологий для разработки программного обеспечения для анализа публикационной активности пользователей в социальной сети “ВКонтакте” с применением элементов геостатистики. Основное внимание уделено архитектуре решения, основанной на технологиях HTML и фреймворке Django, с интеграцией библиотек Folium и spaCy. Рассмотрены механизмы взаимодействия с VK API для получения данных через метод newsfeed.search, а также методы обработки и визуализации полученных данных. Практическое применение работы заключается в возможности отслеживания географического распространения тематических публикаций и выявления локальных центров активности. Результаты исследования можно применять в социальных исследованиях, маркетинговом анализе и мониторинге общественного мнения в социальных сетях.

Ключевые слова: анализ публикационной активности, геостатистика, социальная сеть ВКонтакте, API, Django, folium, spacy, визуализация данных.

Введение.

В современном информационном обществе социальные сети стали ключевым инструментом коммуникации и распространения информации, что обуславливает актуальность исследования публикационной активности пользователей. Всё большую значимость приобретает анализ географических аспектов распространения контента, позволяющий выявлять региональные особенности информационного обмена и определять центры влияния. Несмотря на значительное количество исследований в области анализа социальных медиа, существует потребность в разработке специализированных программных решений, интегрирующих возможности

геостатистического анализа с функционалом сбора и обработки данных из социальных сетей.

Цели и задачи разработки.

Целью данного исследования является разработка программного обеспечения для анализа публикационной активности в социальной сети ВКонтакте с использованием элементов геостатистики. Для достижения поставленной цели определены следующие задачи:

1. Сбор данных: использование VK API для получения данных о публикациях пользователей.

2. Обработка текста: применение библиотеки spaCy для анализа текстового содержимого публикаций.

3. Геостатистический анализ: использование библиотеки Folium для визуализации данных на карте.

4. Создание веб-интерфейса: разработка удобного интерфейса с помощью Django и HTML.

Сбор данных.

Для сбора данных о публикациях пользователей был использован метод запроса newsfeed.search API ВКонтакте. Этот метод позволяет извлекать записи из новостной ленты по заданным параметрам, по таким как ключевые слова, геолокация и временные промежутки.

Пример запроса может выглядеть следующим образом:

```
import requests

def get_vk_posts(query, count=100):
    access_token = 'YOUR_ACCESS_TOKEN'
    url = 'https://api.vk.com/method/newsfeed.search'
    params = {
        'q': query,
        'count': count,
        'access_token': access_token,
        'v': '5.131'
    }
    response = requests.get(url, params=params)
    return response.json()
```

YOUR_ACCESS_TOKEN – токен доступа пользователя либо приложения

В данном примере показан запрос по ключевому слову с определенным количеством постов, по умолчанию, которое равное 1, максимальное значения запрашиваемых постов за один запрос равно 200 постам

Обработка данных.

После получения данных о публикациях с использованием библиотекой spaCy, производится обработка данных. Это включает в себя проверку по семантическому сходству ключевого слова и текста полученной новости, после выполнения которого сортируем новости от наибольшего схожего к наименшему:

```
import spacy
nlp = spacy.load('ru_core_news_sm')
def process_posts(posts):
    query_doc = nlp(query)
    posts_with_similarity = []
    for post in posts:
        if 'text' in post:
            post_doc = nlp(post['text'])
            similarity = query_doc.similarity(post_doc)
            posts_with_similarity.append((post, similarity))
    posts_with_similarity.sort(key=lambda x: x[1], reverse=True)
    posts = [post for post, _ in posts_with_similarity]
    return(posts)
```

Визуализация данных.

Для визуализации результатов анализа на карте используется библиотека Folium. Folium — это мощная библиотека визуализации данных в Python, которая была создана в первую очередь для того, чтобы помочь людям визуализировать геопространственные данные.

```
import folium
def create_map(posts):
    m = folium.Map(location=[55.7558, 37.6173], zoom_start=5) # Центр
    Moskvy
```

for post in posts:

if 'geo' in post:

place = post['geo']['place']

coordinates.append([place['latitude'], place['longitude']])

folium.Marker(

location=[place['latitude'], place['longitude']],

popup=[place['title']],

icon=folium.Icon(color='blue'),

).add_to(m)

return m

При этом есть ограничения. Многие посты в ВКонтаке не сопровождаются геоданными, в следствии чего многие посты не указываются на карту. Что в свою очередь приводит к менее точному анализу географической информации найденных постов.

Создание веб-интерфейса.

Используя фреймворк Django, мы создадим веб-приложение с простым интерфейсом для загрузки данных и отображения результатов анализа.

Пример структуры приложения:

- `views.py`: Логика обработки запросов и возврат результатов.
- `templates/`: HTML-шаблоны для отображения интерфейса.

Результаты.

Разработанное программное обеспечение позволяет пользователям:

1. Выполнять поиск публикаций по заданным ключевым словам.
2. Анализировать данных публикаций с целью выявления геоданных.
3. Визуализировать результаты анализа на интерактивной карте.

Эти функции могут быть полезны для маркетологов, исследователей и социологов, желающих понять географические тренды в публикационной активности пользователей. А созданный интерфейс позволяет интуитивно понять функциональность веб-приложения.

Заключение.

Создание программного обеспечения для анализа публикационной активности в социальной сети с использованием элементов геостатистики

представляет собой актуальную задачу в условиях современного информационного общества. Применение фреймворка Django и библиотек Folium и spaCy позволяет эффективно собирать, обрабатывать и визуализировать данные, что открывает новые возможности для исследования социальных взаимодействий в цифровом пространстве.

В будущем планируется расширение функционала программы, улучшение алгоритмов обработки текста для повышения точности анализа, улучшение интерфейса веб-приложения, чтобы он стал интуитивно понятным и удобным в использовании

Список используемой литературы

1. Воробьев, А. В. Подход к динамической визуализации разнородных геопространственных векторных изображений / А. В. Воробьев, Г. Р. Воробьева // Компьютерная оптика. – 2024. – Т. 48, № 1. – С. 123-138. – DOI 10.18287/2412-6179-CO-1279.
2. Воробьев, А. В. Индуктивный метод восстановления временных рядов геомагнитных данных / А. В. Воробьев, Г. Р. Воробьева // Труды СПИИРАН. – 2018. – № 2(57). – С. 104-133. – DOI 10.15622/sp.57.5.
3. Воробьев, А. В. Концепция единого пространства геомагнитных данных / А. В. Воробьев, Г. Р. Воробьева, Н. И. Юсупова // Труды СПИИРАН. – 2019. – Т. 18, № 2. – С. 390-415. – DOI 10.15622/sp.18.2.390-415.
4. An Approach to Diagnostics of Geomagnetically Induced Currents Based on Ground Magnetometers Data / A. Vorobev, A. Soloviev, V. Pilipenko [et al.] // Applied Sciences (Switzerland). – 2022. – Vol. 12, No. 3. – DOI 10.3390/app12031522.
5. Воробьев, А. В. Подход к обнаружению и устраниению артефактов пространственных изолиний в приложениях Веб-ГИС / А. В. Воробьев, Г. Р. Воробьева // Компьютерная оптика. – 2023. – Т. 47, № 1. – С. 126-136. – DOI 10.18287/2412-6179-CO-1127.
6. ВКонтакте API Documentation: <https://vk.com/dev/manuals>
7. spaCy Documentation: <https://spacy.io/usage>
8. Folium Documentation: <https://python-visualization.github.io/folium/>