

**«Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И.Ульянова (Ленина)»  
(СПбГЭТУ «ЛЭТИ»)**

<b>Направление</b>	09.03.04 – Программная инженерия
<b>Профиль</b>	Разработка программно-информационных систем
<b>Факультет</b>	КТИ
<b>Кафедра</b>	МО ЭВМ

*К защите допустить*

Зав. кафедрой

Кринкин К.В.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
БАКАЛАВРА**

**ТЕМА: РАЗРАБОТКА МОДУЛЯ АВТОМАТИЗАЦИИ ИМПОРТА И ГЕО-  
КОНТЕКСТНОЙ РАЗМЕТКИ ОТКРЫТЫХ ДАННЫХ**

Студент	<hr/>	Блеес Э.И.
	<i>подпись</i>	
Руководитель	<hr/>	Заславский М.М.
<i>(Уч. степень, уч. звание)</i>	<i>подпись</i>	
Консультанты	<hr/>	Иванов А.Н.
<i>(Уч. степень, уч. звание)</i>	<i>подпись</i>	
	<hr/>	Яценко И.В.
<i>(Уч. степень, уч. звание)</i>	<i>подпись</i>	

Санкт-Петербург

2017

## ЗАДАНИЕ

Утверждаю

Зав. кафедрой МО ЭВМ

\_\_\_\_\_ Кринкин К.В.

«        » 20 \_\_\_\_ г.

Студент            Блеес Э.И.

Группа 3304

Тема работы: Разработка модуля автоматизации импорта и геоконтекстной разметки открытых данных

Место выполнения ВКР: СПбГЭТУ «ЛЭТИ», кафедра МО ЭВМ

Исходные данные (технические требования):

Документация платформы Geo2Tag, источники открытых данных (официальный сайт открытых данных Санкт-Петербурга, социальная сеть «ВКонтакте»)

## Содержание ВКР:

Введение, обзор предметной области, выбор метода решения, описание метода решения, исследование свойств решения, заключение

Перечень отчетных материалов: пояснительная записка, иллюстративный материал

## Дополнительные разделы: Безопасность жизнедеятельности

Дата выдачи задания

Дата представления ВКР к защите

«24 » апреля 2017 г.

«15 » июня 2017 г.

Студент

Блеес Э.И.

подпись

## Руководитель

Заславский М.М.

(Уч. степень, уч. звание)

подпись

## Консультанты

Иванов А.Н.

(Уч. степень, уч. звание)

подпись

Яценко И.В.

(Уч. степень, уч. звание)

подпись

## КАЛЕНДАРНЫЙ ПЛАН ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Утверждаю

Зав. кафедрой МО ЭВМ

Кринкин К.В.

« » 20 г.

Студент            Блеес Э.И.

Группа 3304

Тема работы: Разработка модуля автоматизации импорта и геоконтекстной разметки открытых данных

№ п/п	Наименование работ	Срок вы- полнения
1	Обзор литературы по теме работы	24.04 – 30.04
2	Наименование раздела	01.05 – 04.05
3	Наименование раздела	05.05 – 19.05
4	Наименование раздела	20.05 – 24.05
5	Предзащита	30.05
6	Оформление пояснительной записки	25.05 – 01.06
7	Оформление иллюстративного материала	27.05 – 15.06

Студент

Блеес Э.И.

## Руководитель

Заславский М.М.

(Уч. степень, уч. звание)

## РЕФЕРАТ

Выпускная квалификационная работа содержит пояснительную записку объёмом 77 страниц, 31 рисунок, 72 источника информации.

Ключевые слова: открытые данные, LBS, геоконтекст, Geo2Tag.

Объект исследования: открытые данные.

Работа посвящена исследованию проблемы отсутствия интерфейса, позволяющего получить для обработки открытые данные, а также отсутствия инструментов импорта и обработки этих данных.

Актуальность проблемы связана с быстрым ростом количества открытых данных, а также активным внедрением технологий контекстно-зависимых вычислений.

В работе дан подробный обзор предметной области, а также существующих решений, в том числе были рассмотрены Big Data решения, ГИС, а также LBS-платформы. LBS-платформа Geo2Tag оказалась наиболее подходящей для решения поставленной задачи.

Приводится решение проблемы путем создания плагина к открытой LBS-платформе Geo2Tag, решающего следующие подзадачи:

- Получение данных из двух источников – сайта открытых данных Санкт-Петербурга и социальной сети «ВКонтакте»;
- Обработка полученных данных. Выделение дополнительного признака данных, основываясь на числовом показателе, используя методы расчёта среднего арифметического, медианы набора данных, а также метод кластеризации;
- Добавление обработанных данных в базу LBS-платформы Geo2Tag.

Приведены и проанализированы результаты исследования полученного решения: время работы плагина не зависит от промежутка времени выборки, но линейно зависит от количества анализируемых точек Wi-Fi.

## **ABSTRACT**

The paper provides a detailed overview of the subject area, as well as existing solutions, including consideration of Big Data solutions, GIS and LBS-platforms. LBS-platform Geo2Tag was the most suitable for solving the task.

The solution of the problem is provided by creating a plug-in to the open LBS-platform Geo2Tag, which solves the following subtasks:

- Receiving data from two sources – the site of open data of St. Petersburg and VKontakte social network;
- Processing of the received data. The allocation of an additional characteristic of the data, based on the numerical characteristic, using cluster analysis, mean and median of numeric data;
- Importing processed data to the database of the Geo2Tag LBS platform.

Results of the study of the obtained solution are presented and analyzed: the operating time of the plugin does not depend on the time interval, but depends linearly on the number of analyzed Wi-Fi points.

## Содержание

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.....	8
ВВЕДЕНИЕ.....	9
1. ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ.....	11
1.1. Открытые данные .....	11
1.2. Контекстно-зависимые вычисления .....	14
1.3. Геоконтекстная разметка .....	15
1.4. LBS-платформы .....	16
1.5. LBS-платформа Geo2Tag .....	17
1.6. Анализ аналогов.....	17
1.7. Выводы .....	19
2. ВЫБОР МЕТОДА РЕШЕНИЯ .....	20
2.1. Выбор функций обработки данных .....	20
2.2. Выбор источников данных .....	22
2.3. Выводы .....	23
3. ОПИСАНИЕ МЕТОДА РЕШЕНИЯ .....	24
3.1. Архитектура плагина.....	24
3.2. Интерфейсы плагина импорта открытых данных.....	25
3.3. Описание алгоритмов работы .....	25
3.3.1. Получение данных .....	26
3.3.2. Обработка данных.....	28
3.3.3. Добавление данных в базу.....	31
3.4. Тестирование решения.....	31
3.4.1. Юнит-тестирование.....	31
3.4.2. Интеграционное тестирование.....	32
3.5. Выводы .....	33
4. ИССЛЕДОВАНИЕ СВОЙСТВ РЕШЕНИЯ .....	34
4.1. Условия эксперимента .....	34

4.2. Используемые инструменты .....	34
4.3. Результаты эксперимента .....	36
4.3.1. Зависимость времени работы плагина от промежутка времени выборки .....	39
4.3.2. Зависимость времени работы плагина от количества анализируемых точек Wi-Fi .....	46
4.4. Выводы .....	54
5. БЕЗОПАСНОСТЬ ЖИЗНЕДЕЯТЕЛЬНОСТИ .....	55
5.1. Приемлемость организации диалога для выполнения поставленных задач .....	55
5.2. Информативность .....	57
5.3. Соответствие ожиданиям пользователя .....	58
5.4. Пригодность для обучения .....	58
5.5. Контролируемость .....	59
5.6. Устойчивость к ошибкам .....	60
5.7. Адаптируемость к индивидуальным особенностям пользователя .....	61
5.8. Выводы .....	61
ЗАКЛЮЧЕНИЕ .....	62
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	63
Приложение А_Схема архитектуры плагина .....	69
Приложение Б_Экспериментально полученные данные .....	70

## **ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ**

В настоящей пояснительной записке применяют следующие термины с соответствующими определениями:

БД – База данных

ГИС – Географическая информационная система

ОС – Операционная система

API – Application programming interface

CSV – Comma-separated values

HTML – Hypertext Markup Language

JSON – JavaScript Object Notation

LBS – Location-based service

MS – Microsoft

ODF – Open Document Format

OOXML – Office Open XML

OS – Operating system

PDF – Portable Document Format

RDF – Resource Description Framework

REST – Representational state transfer

XML – Extensible Markup Language



# **ВВЕДЕНИЕ**

## **Актуальность.**

Проблема отсутствия удобного для исследователя интерфейса для работы с открытыми данными, а также отсутствия инструментов импорта и обработки геоконтекстных данных актуальна в связи с быстрым ростом объема открытых данных [1] и активного внедрения технологий контекстных вычислений [2]. Согласно прогнозу на 2019 год, этот фактор приведет к серьезному росту рынка LBS за счет интеграции технологии контекстной разметки с существующими технологиями LBS-платформ [3]. Использовать готовое решение (LBS платформы), вместо написания своей реализации - более выгодно и быстро.

## **Цель работы.**

Разработать модуль автоматизации импорта и геоконтекстной разметки открытых данных в LBS платформу.

## **Постановка задачи.**

Для достижения цели необходимо решить следующие задачи:

- Исследовать существующие решения по обработке открытых данных, решения по работе с геоданными;
- Провести обзор LBS-платформ;
- Спроектировать архитектуру плагина LBS-платформы, в том числе решить следующие подзадачи:
  - Получение данные;
  - Обработка полученных данных;
  - Добавление обработанных данных в базу;
- Провести тестирование плагина;
- Исследовать производительность полученного решения.

## **Объект исследования.**

Открытые данные.

**Предмет исследования.**

Автоматизация импорта и геоконтекстной разметки открытых данных.

**Практическая значимость.**

Плагин автоматизации импорта и геоконтекстной разметки открытых данных позволит улучшить импорт открытых данных в LBS платформе Geo2Tag, дав возможность клиентам данной платформы работать с обработанными открытыми данными.

**Опубликованные работы по теме.**

1. Разработка модуля автоматизации импорта и геоконтекстной разметки открытых данных / Э.И. Блеес, М.М. Заславский // 70-я Научно-техническая конференция профессорско-преподавательского состава университета: сборник докл. студентов, аспирантов и молодых ученых. - 2017. - С. 115-118

## 1. ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

Перед тем, как начать решать поставленную задачу, необходимо определить основные понятия, а также исследовать актуальные работы по данному вопросу.

### 1.1. Открытые данные

Открытые данные — концепция, отражающая идею о том, что определённые данные должны быть свободно доступны для машиночитаемого использования и распространения без ограничений авторского права, патентов и других механизмов контроля [4]. Цели движения открытых данных похожи на цели движения за открытое программное обеспечение (open source) [5] и за открытый доступ (open access) [6]. Идея открытости данных обрела популярность с распространением интернета и, прежде всего, с запуском правительственных инициатив Data.gov [7] и Data.gov.uk [8].

Контроль доступа и использования данных осуществляется государством и частными организациями через ограничения, лицензии, копирайт, патенты и требования оплаты для доступа или повторного использования. Сторонники идеи «открытых данных» считают, что подобные ограничения идут против общественного блага и данные должны быть доступны без ограничений или оплаты. Также важно, что данные должны быть доступны без последующих запросов на разрешение, хотя и способы повторного использования, такие как создание продуктов на базе данных, могут контролироваться лицензией.

Объем открытых данных огромен, и будет только увеличиваться. На данный момент их объем составляет 4.4 зеттабайта, т.е. 4.4 триллиона Гб. К 2020 году объем открытых данных увеличится десятикратно и составит 44 зеттабайта [1].

Работа с открытыми данными усложняется из-за разнообразия форматов их хранения. Краткая характеристика форматов хранения открытых данных:

## **JSON**

JSON — текстовый формат обмена данными, легко читаемый людьми, а также легко обрабатываемый машиной. За счет своей лаконичности успешно используется для обмена данными, а также для сериализации структур данных. Стандарт [9].

## **XML**

XML — расширяемый язык разметки. Широко используется для обмена данными, но часто бывает избыточен (в целях разработки языка прямо указано, что лаконичность не является приоритетом проекта). В описании документов, где доля разнотипных символьных данных велика, а доля разметки мала — XML успешен. Стандарт [10].

## **RDF**

RDF — модель представления данных, разработанная консорциумом Всемирной паутины W3C [11]. Используется для оформления баз знаний или представления их данных в рамках различных проектов в множестве институтов, университетов и иных организаций. В основном, в научно-исследовательских целях. RDF сам по себе является не форматом файла, а абстрактной моделью данных [12], то есть описывает предлагаемую структуру, способы обработки и интерпретации данных. Для хранения и передачи информации, уложенной в модель RDF, существует целый ряд форматов записи. Стандарт [13].

## **Электронные таблицы**

Электронная таблица — компьютерная программа, позволяющая проводить вычисления с данными, представленными в виде двумерных массивов, имитирующих бумажные таблицы [14]. Примерами электронных таблиц являются Microsoft Excel [15], LibreOfficeCalc [16] и другие. Формат хранения напрямую зависит от программы, поэтому обрабатывать такие данные затруднительно.

### **Файлы с разделителями**

CSV — текстовый формат, предназначенный для представления табличных данных, в котором каждая строка файла — это одна строка таблицы. Разделителем значений колонок является символ запятой, однако на практике часто используются другие разделители. Стандарт [17].

### **Текстовый документ**

Для демонстрации многих типов данных может быть достаточно форматов документов типа doc [18], ODF [19], OOXML [20] или PDF [21].

### **Простой текст**

Документы в формате .txt легко читаются машиной, однако, структура данных, как правило, отсутствует. При обмене текстовыми файлами между операционными системами возможны проблемы. MS Windows, Mac OS и прочие Unix-подобные ОС по-разному сообщают компьютеру, что достигли конца строки.

### **Сканы документов**

Наименее подходящая форма для большинства данных. Может пригодиться для отображения данных, которые не были рождены в электронной форме (типичным примером может послужить архивный материал) — картинка лучше, чем ничего.

### **HTML**

HTML — стандартизированный язык разметки документов в Web. Большинство веб-страниц содержат описание разметки на языке HTML. Язык HTML интерпретируется браузерами; полученный в результате интерпретации форматированный текст отображается на экране монитора компьютера или мобильного устройства. Стандарт [22].

Самыми удобными для работы форматами данных являются JSON и XML, т.к. они структурированно хранят информацию об объекте, а также потому что существует множество инструментов для работы с этими форматами данных.

## 1.2. Контекстно-зависимые вычисления

Контекстно-зависимые вычисления изучают контекстно-зависимые системы (аппаратное и программное обеспечение), анализирующие состояние пользователя и окружающую среду, и адаптирующие свою работу под изменяющиеся условия.

Основными источниками информации для контекстно-зависимых систем являются местоположение, социальное и физическое окружение. Несмотря на то, что определение местоположения широко используется в настоящее время, оно не всегда учитывает изменяющиеся интересы пользователя. Контекстная зависимость в широком понимании включает в себя находящиеся рядом людей, устройства, доступ к интернету, уровень освещенности, уровень шума, а также взаимодействие людей в бытовых ситуациях. Например, находитесь ли вы сейчас с семьей или со своим школьным другом.

В настоящее время исследователями используются различные определения контекста. Профессор университета Карнеги Меллон, Анинд К. Дей [2] дает такое определение: «Контекст — это любая информация, использующаяся для описания ситуации или объекта». В данном случае под объектом подразумевается место или человек, которые могут взаимодействовать между пользователем или приложением. Также контекст может рассматриваться как любое окружение объекта, включая сам объект и его взаимодействия со средой. Если какая-либо информация может быть использована для определения и описания события или взаимодействия объекта, тогда можно назвать её контекстными характеристиками. Однако во многих случаях исследователи используют то определение контекста, которое, по их собственному мнению, подходит для отражения основного смысла. Иногда такие определения представляют контекст в общих чертах. Билл Склилит и Марвин Теймор [23], ученые из Xerox PARC, определили контекст как местоположение, социальное окружение, объекты и изменения этих объектов. Альбрехт Скимидт [24-25]

рассматривает контекст как знания о состоянии пользователя и устройства, включая окружающую среду и социальное окружение.

### **1.3. Геоконтекстная разметка**

Геоконтекст – это любая информация о местоположении сущности, которая может описать её текущее состояние [26].

В статье «A Survey of People Movement Analytics Studies in the Context of Smart Cities» [27] было дано полное описание задач и решений, связанных со сбором, хранением, анализированием, визуализированием, использованием и распространением данных о перемещении людей, и объединением подобных данных в контексте умного города и интернета вещей. Можно выделить следующие актуальные задачи:

- Indoor-навигация. Отслеживание на "открытой местности" уже показывает отличную точность и имеет множество данных для анализа. Перемещения внутри помещений сложнее для анализа в связи с малым количеством данных;
- 3D анализ. 2D по широте/долготе широко распространен. А анализ 3D, включая высоту очень сложен. Особенно актуально для анализа перемещений в здании, учитывая перемещения по этажам и прочее;
- Онтологический подход. Использование как и данных о траекториях, так и вычисленные шаблоны траекторий;
- Проблемы приватности данных о местоположении;
- Карты внутри помещений;
- Платформы открытого доступа. Для лучшего вовлечения жителей в создание данных, необходимо создавать новые платформы, доступные для использования.

Полезным для получения геоконтекста данных методом является геокодирование. Геокодирование – это процесс преобразования адресов (например,

"1600 Amphitheatre Parkway, Mountain View, CA") в географические координаты (например, широта 37,423021 и долгота -122,083739), которые можно использовать для размещения маркеров на карте или ее позиционирования [28].

По содержанию в открытых данных геоконтекстной информации их можно разделить на 3 типа:

- Данные, не содержащие в структуре никакой геоконтекстной информации;
- Данные, содержащие в структуре скудную информацию о расположении сущности. Например страна, город;
- Данные, содержащие в структуре полный адрес расположения сущности;
- Данные, содержащие в структуре географические координаты сущности.

Последний тип данных идеально подходит для анализа, т.к. очень легко может быть визуализирован на карте, не выполняя каких-либо других вычислений.

При выборе источника данных для работы, приоритетом будет последний тип данных.

#### **1.4. LBS-платформы**

Location-based service (LBS) — программный сервис, использующий данные о локации для управления какими-либо функциями [29]. LBS может быть использован для многих сфер, например, для indoor-навигации, контроля физической активности и т.д.

Некоторые LBS-платформы собирают данные с множества мобильных устройств и предоставляют разработчикам SDK [30]. Также существуют LBS-платформы, которые дают пользователю возможность хранить и обрабатывать



импортированные данные. При любом подходе LBS-платформы облегчают разработку приложений, использующих геоконтекст.

### **1.5. LBS-платформа Geo2Tag**

Geo2tag это LBS-платформа с открытым исходным кодом. Платформа предоставляет доступ ко всем программным интерфейсам и сервисам (база геоданных, фильтры, запросы), необходимым для развертывания геоконтекстных сервисов, основанных на местоположении. Предоставляет пользователю следующую функциональность:

- Пространственно-временное хранение данных;
- Пространственно-временная обработка данных;
- Изоляция данных внутри выделенных сервисов.

В основе Geo2Tag лежит Python-фреймворк Flask [31], использующий нереляционную БД MongoDB [32] и веб-сервер Apache [33]. Все взаимодействия с платформой, включая администрирование и создание новых геосервисов, осуществляются при помощи интерфейсов прикладного программирования, отвечающих стандарту REST [34-35].

### **1.6. Анализ аналогов**

В связи с огромным количеством открытых данных [1] стоит рассматривать решения, специализирующиеся на обработке больших объемов данных. Big Data — совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence [36-37]. Примеры Big Data решений: Hadoop [38], Spark [39], Storm [40], Teradata [41], Cassandra [42], Oracle Big Data [43].

Говоря об обработке открытых данных, в том числе геоданных, нельзя ни упомянуть ГИС. Геоинформационная система — система сбора, хранения, анализа и графической визуализации географических данных и связанной с ними информации о необходимых объектах. Понятие геоинформационной системы также используется в более узком смысле — как программного продукта, позволяющего пользователям искать, анализировать и редактировать как цифровую карту местности, так и дополнительную информацию об объектах [44]. Примерами таких систем являются различные геобазы данных, такие как ArcGIS [45]. К ним же можно отнести инструменты GIS as a service [46] — Google Maps [47], Apple Maps [48], Bing Maps [49], OpenStreetMap [50], CartoDB [51], MapBox [52].

В явном виде не существует интерфейсов для импорта в LBS [35]. Инструменты для работы с BigData не подходят, т.к. это инструмент более широкого применения, в основном используемый для хранения и обработки постоянно и быстро накапливающихся данных. Геобазы данных в основном предоставляют API, которые удобны для прикладных программистов, но для исследователей/ученых необходим более высокоуровневый инструмент.

Рассмотрим другие LBS платформы — Anagog [30], Azoft LBS-platform [53], TomTom [54]. Anagog и TomTom предоставляют SDK, собирающую данные с множества мобильных устройств, в том числе данные GPS, датчиков. Данные поступают постоянно, и решения, использующие SDK, постоянно их обрабатывают. Azoft LBS-platform — закрытая платформа, собирающая данные с мобильных устройств и предоставляющая готовые решения для очевидных сценариев использования.

Сравнение LBS-платформ представлено в табл. 1.

Таблица 1 – Сравнение LBS-платформ

	Возможность сохранения данных для анализа	Возможность создавать пла- гины к плат- форме	Открытость платформы	Наличие SDK	Возможность получения и обработки дан- ных в реальном времени	Получение готовых ре- шений
Anagog	Нет	Нет	Нет	Да	Да	Нет
TomTom	Нет	Нет	Нет	Да	Да	Нет
AzoftLBS	Нет	Нет	Нет	Нет	Нет	Да
Geo2Tag	Да	Да	Да	Нет	Нет	Нет

Как видно из сравнения, в основном LBS платформы направлены на выдачу и обработку данных в реальном времени, что связывает их с Big Data решениями. Для выбранной цели подходит лишь LBS-платформа Geo2Tag, т.к. она открытая и позволяет создавать к ней плагины, а также предоставляет возможность сохранения данных для последующей обработки.

### 1.7. Выводы

В главе были определены основные понятия, а также исследованы актуальные работы по данному вопросу. Был проведен обзор Big Data решений, ГИС, а также LBS-платформ. LBS-платформа Geo2Tag оказалась наиболее подходящей для решения поставленной задачи.

## **2. ВЫБОР МЕТОДА РЕШЕНИЯ**

Для решения поставленной задачи принято решение разработать плагин к LBS-платформе Geo2Tag так как:

- Инструмент обработки и импорта данных должен решать задачи обработки открытых данных и импорта их в базу;
- Платформа Geo2Tag открытая и расширяемая;
- В предыдущих работах [35] были разработаны интерфейсы для плагинов импорта данных в LBS платформу Geo2Tag. Описываемое решение предоставляет интерфейс для импорта широкого класса источников, но не предоставляет никаких инструментов обработки.

### **2.1. Выбор функций обработки данных**

Необходимо определить, какие функции обработки данных должен иметь плагин. Для этого был проведен обзор методов и алгоритмов извлечения геоконтекста из открытых данных.

Отбор источников для обзора основывался на следующем критерии: статья (книга) должна содержать подробный обзор технологий и методов обработки геоданных или описывать применение некоторых техник на конкретном примере.

В книге «From GeoVisualization to visual-analytics: methodologies and techniques for human-information discourse» [55] дано подробное описание технологий анализа больших геопространственных наборов данных, а также другие, связанные с обработкой геоданных, темы, такие как:

- технологии, основанные на положении человека и ориентации его мобильного девайса. Например, позволяющие отображать на дисплей мобильного устройства распознанные препятствия, соотнесенные с тем, что "видит" камера;

- технологии, соединяющие геоданные и данные соц.сетей. Пример использования - отображение на карте и на дисплее мобильного устройства (с включенной камерой) местоположение друзей, путей к ним;
- Chorem технологии (территориальные). Позволяют анализировать геоданные, связанные с областями, территориями - например страны, их деление и связанные с ними темы;
- Технологии маркетинга. Пример: пользователь включает камеру и "наводит" на витрину магазина, рядом с которым он стоит. Система определяет по местоположению пользователя и ориентации его устройства, о каком магазине идет речь, и выводит информацию, предложенную магазином - например информация о акциях, скидках и т.д.;
- Анализ информации через облако тегов.

Было принято решение реализовать методы извлечения дополнительного признака из набора данных, основываясь на некотором числовом показателе. Идея заключается в вычислении среднего арифметического или медианы числового показателя набора данных, и определении дополнительного признака на основе отношения значения числового показателя к вычисленным характеристикам набора.

В статье «How to study the City on Instagram» [56] было проведено исследование, основанное на обработке огромного количества данных, полученных из социальной сети Instagram [57]. Предметом исследования статьи является "Urban dynamics" или другими словами жизнь города. Исследователи отмечают, что в похожих работах в основном рассматривались места проживания людей, плотность заселения территории. Исследователи же хотят использовать данные из Instagram, чтобы узнать самые популярные места в городе для отдыха, встреч людей. Т.е. где проходит основной досуг людей, а не их рутинная жизнь/работа и т.д. Так же исследователи считают, что есть возможность выделять социальные группы и места их присутствия в городе. Исследователи

анализировали Амстердам и Копенгаген. В исследовании использовались различные методы анализа данных – анализ связности подграфов, анализ сетей, классификация, кластеризация.

Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы [58]. Соответственно, после кластеризации набора данных, можно выделить дополнительный признак, основываясь на принадлежности элемента набора к определённому кластеру. В связи с этим было принято решение реализовать метод кластеризации набора данных, основываясь на числовом показателе, для извлечения дополнительного признака в наборе данных.

## **2.2. Выбор источников данных**

Как было указано в разделе 1 – приоритетом при поиске источников открытых данных будут являться данные, хранимые в формате JSON или XML, из-за структурированности данных форматов, а также наличие множества инструментов для удобной работы с ними. Также данные должны содержать гео-контекстную информацию, желательно географические координаты объекта.

Был найден подходящий источник – набор данных «Бесплатные зоны Wi-Fi в общественных местах в Санкт-Петербурге», расположенный на сайте открытых данных Санкт-Петербурга [59]. В структуре данных присутствуют географические координаты и адресный ориентир, что позволит позиционировать данные на карте. Также присутствует информация о величине зоны покрытия, что позволяет точно знать масштабы зоны Wi-Fi.

Необходим какой-либо числовой показатель, который позволит, используя выбранные методы обработки, выделить дополнительный признак данных. Так как известны координаты точки и зона её покрытия, принято решение получить данные о количестве записей, опубликованных в социальной сети «ВКонтакте» в определенный промежуток времени в области действия точки [60].

Формат данных, получаемых из обоих источников – JSON. Так как данные в Geo2Tag тоже хранятся в этом формате, усилия на их обработку будут заметно ниже, в сравнении с другими форматами, в том числе и XML.

### **2.3. Выводы**

В данной главе были выбраны методы обработки данных, которые необходимо реализовать в плагине:

- Вычисление среднего арифметического числового показателя набора данных, и определения дополнительного признака на основе отношения значения числового показателя к вычисленным характеристикам набора;
- Вычисление медианы числового показателя набора данных, и определения дополнительного признака на основе отношения значения числового показателя к вычисленным характеристикам набора;
- Кластеризация набора данных, основываясь на числовом показателе, для извлечения дополнительного признака в наборе данных.

Были выбраны источники открытых данных:

- Набор данных «Бесплатные зоны Wi-Fi в общественных местах в Санкт-Петербурге», расположенный на сайте открытых данных Санкт-Петербурга;
- Социальная сеть «ВКонтакте», а именно записи, опубликованные в сети в определенный промежуток времени в области действия точки Wi-Fi.

В итоге плагин должен решать следующую общую задачу: сопоставление двух массивов данных для выделения дополнительного признака из первого массива и импорт данных в базу.

### 3. ОПИСАНИЕ МЕТОДА РЕШЕНИЯ

В данной главе дано подробное описание плагина импорта и геоконтекстной разметки открытых данных – описана архитектура решения, алгоритмы получения, обработки данных и добавления их в базу.

#### 3.1. Архитектура плагина

Схема архитектуры плагина изображена на рис. 1. Увеличенный вариант схемы находится в приложении А.

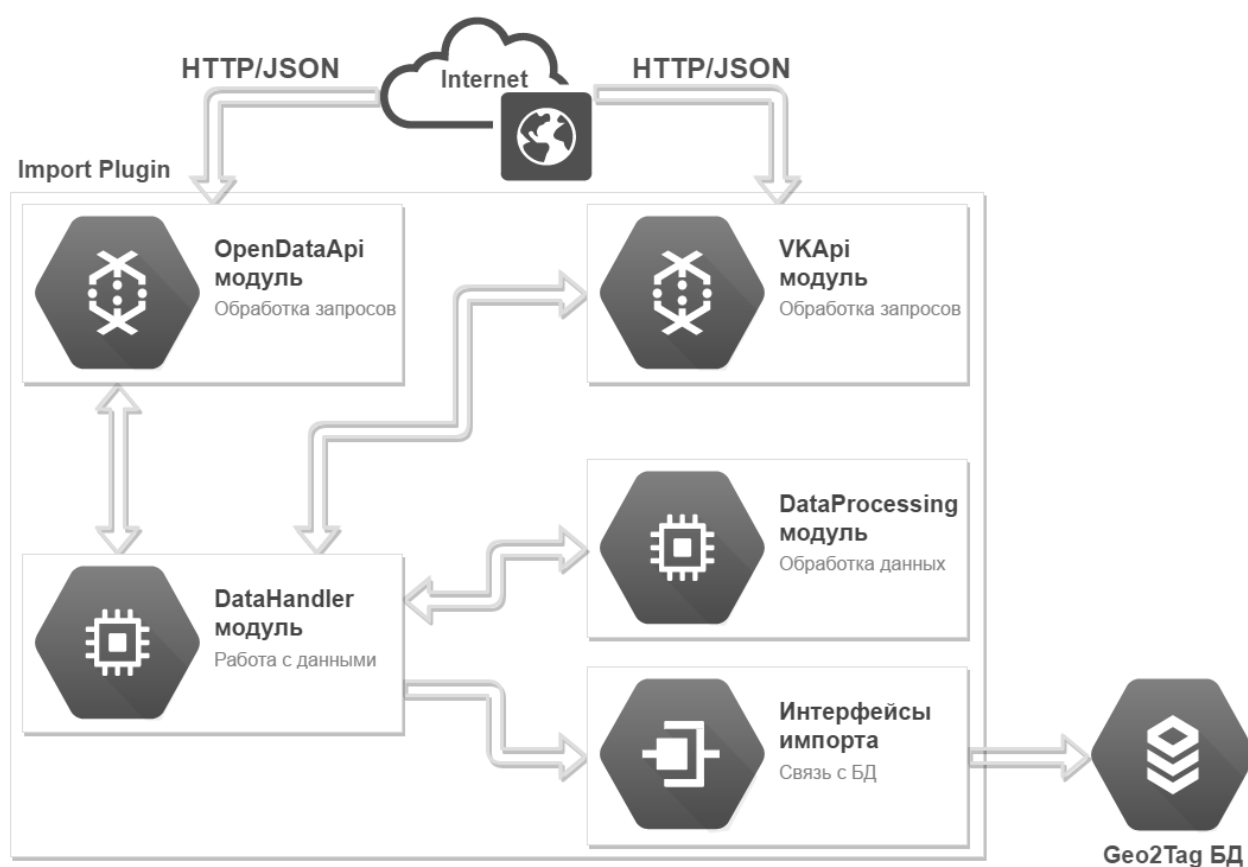


Рисунок 1 – Архитектура плагина

Модули **OpenDataApi** и **VKApi** ответственны за получение данных из источников – выполнение запросов и обработка ответов. В модуле **DataHandler** выполняется основная работа – получение исходных данных от модулей загрузки (**OpenDataApi** и **VKApi**), обработка, используя методы, реализованные в модуле **DataProcessing**, передача обработанных данных в интерфейсы импорта. Интерфейсы импорта – реализованные интерфейсы импорта данных в платформу **Geo2Tag**.



Как было указано в разделе 1.5 – LBS-платформа Geo2Tag основана на Python-фреймворке Flask, и большая часть кода написана на языке Python. В связи с этим плагин также был написан на языке Python.

Плагин работает в асинхронном режиме, поэтому не блокирует работу пользователя во время выполнения своей задачи.

### **3.2. Интерфейсы плагина импорта открытых данных**

Как уже было отмечено ранее, проблема отсутствия интерфейсов для импорта данных в LBS-платформу была сформулирована в статье [35], авторы которой разработали интерфейсы для плагинов импорта данных в LBS платформу Geo2Tag. Так как в разработанном плагине используются данные интерфейсы, необходимо дать их описание.

Алгоритм импорта данных состоит из четырех стадий:

- Получение данных – установление соединения, авторизация и загрузка данных от источника открытых данных;
- Разделение массива открытых данных на отдельные элементы, десериализация массива данных;
- Трансляция отдельных элементов открытых данных в формат Geo2Tag;
- Запись точек в БД сервиса.

Интерфейсы импорта открытых данных включают в себя логику выполнения каждого этапа алгоритма импорта.

В связи с нетривиальностью поставленной задачи (два источника, анализ данных) получение и обработка данных плагином производится в разработанных модулях вне интерфейсов импорта.

### **3.3. Описание алгоритмов работы**

В этом разделе дано подробное описание алгоритмов получения, обработки данных и добавления их в базу.

### 3.3.1. Получение данных

Плагин запускается через REST-интерфейс со следующими параметрами:

- `ChannelName` – название канала, в который будут добавляться точки. Тип данных - `string`;
- `Days_to_load` – количество дней, образующих интервал от даты запуска плагина, за который будут загружены данные о записях, опубликованных в социальной сети «ВКонтакте». Тип данных - `int`;
- `Points_to_load` – количество анализируемых точек Wi-Fi. Тип данных - `int`;
- `Method` – используемый метод анализа данных; Тип данных - `string`.

Как было сказано выше, в модуле `DataHandler` реализована основная логика работы плагина. Модуль на вход получает параметры работы – количество точек Wi-Fi для загрузки, количество дней, определяемое интервал, за который запрашиваются записи в социальной сети «ВКонтакте», а также метод обработки данных, который необходимо использовать.

Первый шаг – получение открытого набора данных «Бесплатные зоны Wi-Fi в общественных местах в Санкт-Петербурге». Для этого используется модуль `OpenDataApi`, в котором реализована работа с API сайта открытых данных Санкт-Петербурга, в том числе метод получения набора данных по его идентификационному номеру. В разделе 2.2 уже упоминалось, что данные, предоставляемые сайтом открытых данных Санкт-Петербурга, имеют формат JSON. Работа с этим форматом в Python велась с помощью модуля `json` [61], поэтому информации из данных не требует дополнительных усилий и какой-либо дополнительной обработки. Схема JSON-объектов из данного набора данных представлена на рис. 2.

```

{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "definitions": {},
  "properties": {
    "Free Wi-Fi": {
      "properties": {
        "address": {
          "type": "string"
        },
        "coverage": {
          "type": "integer"
        },
        "district": {
          "type": "string"
        },
        "latitude": {
          "type": "number"
        },
        "longitude": {
          "type": "number"
        },
        "name_object": {
          "type": "string"
        },
        "name_wifi": {
          "type": "string"
        },
        "note": {
          "type": "string"
        },
        "number": {
          "type": "string"
        },
        "status": {
          "type": "string"
        }
      },
      "type": "object"
    }
  },
  "type": "object"
}

```

Рисунок 2 – Схема JSON-объекта из набора данных «Бесплатные зоны Wi-Fi в общественных местах в Санкт-Петербурге»

Вторым шагом необходимо получить числовой показатель, который будет необходим для анализа данных и выделения дополнительного признака. Так как среди полученных данных о точках Wi-Fi содержатся географические координаты точки и радиус области её действия, есть возможность воспользоваться модулем VKApi, в котором реализован метод получения количества записей из социальной сети «ВКонтакте», прикрепленных к указанной области (географические координаты центра, радиус области), опубликованных в определенный промежуток времени. Данный метод вызывается для каждой

точки. Схема JSON-объекта, представляющего запись в социальной сети «ВКонтакте» представлена на рис. 3.

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "definitions": {},
  "properties": {
    "response": {
      "items": {
        "type": "integer"
      },
      "type": "array"
    }
  },
  "type": "object"
}
```

Рисунок 3 – Схема JSON-объекта, представляющего запись в социальной сети «ВКонтакте»

### 3.3.2. Обработка данных

В модуле `DataHandler` содержится массив данных о точках Wi-Fi в формате JSON. С помощью модуля `json`, к структуре всех JSON-объектов из массива добавляется новое поле – количество записей, полученное из модуля `VKApi`.

Следующий шаг – извлечение дополнительного признака из данных. Весь анализ данных и логика получения дополнительного признака реализованы в модуле `DataProcessing`, задачами которого являются:

- Анализ числового показателя данных;
- Выделение дополнительного признака;
- Формирование структуры данных в JSON формате, хранимой в базе данных `Geo2Tag – point` (точка).

В модуле `DataProcessing` реализовано три метода обработки данных – выделение среднего арифметического, выделение медианы набора, а также кластеризация. Все три метода на вход получают массив точек Wi-Fi, и анализируют количество записей, сделанных в области действия каждой точки.

Методы вычисления среднего арифметического и медианы количества записей реализованы в модуле `DataProcessing`. В зависимости от вычисленного значения среднего или медианы, выделяется дополнительный признак точки

Wi-Fi – её «популярность» – отношение количества записей, опубликованных в социальной сети «ВКонтакте» в области действия точки, к среднему арифметическому или медиане этого показателя по Санкт-Петербургу.

При кластеризации точек по количеству записей используется метод *k*-средних [62], реализованный в модуле *cluster* [63]. Точки делятся на два кластера – популярные (большое количество опубликованных записей) и менее популярные (мало опубликованных записей).

Формируется массив JSON-объектов, соответствующих структуре *point*, хранимой в базе данных *Geo2Tag*. Для этого географические координаты точек Wi-Fi извлекаются из имеющихся объектов, и помещаются в поле «*location*». Вся информация о точке помещается в поле «*json*». Выделенный дополнительный признак «популярность» также входит в поле «*json*», представляющее из себя JSON-объект. Помимо этого, в новой точке фиксируется время импорта. Схема JSON-объекта *Point* представлена на рис. 4.

```

{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "definitions": {},
  "properties": {
    "_id": {
      "properties": {
        "$oid": {
          "type": "string"
        }
      },
      "type": "object"
    },
    "alt": {
      "type": "integer"
    },
    "bc": {
      "type": "boolean"
    },
    "channel_id": {
      "properties": {
        "$oid": {
          "type": "string"
        }
      },
      "type": "object"
    },
    "date": {
      "type": "number"
    },
    "json": {
      "properties": {
        "address": {
          "type": "string"
        },
        "coverage": {
          "type": "string"
        },
        "district": {
          "type": "string"
        },
        "name_object": {
          "type": "string"
        },
        "name_wifi": {
          "type": "string"
        },
        "note": {
          "type": "string"
        },
        "number": {
          "type": "string"
        },
        "popularity": {
          "type": "string"
        },
        "status": {
          "type": "string"
        }
      },
      "type": "object"
    },
    "location": {
      "properties": {
        "coordinates": {
          "items": {
            "type": "number"
          },
          "type": "array"
        },
        "type": {
          "type": "string"
        }
      },
      "type": "object"
    }
  },
  "type": "object"
}

```

Рисунок 4 – Схема JSON-объекта Point

### **3.3.3. Добавление данных в базу**

Готовые к импорту в базу данные возвращаются в модуль `DataHandler`, откуда передаются в модуль импорта `OpenDataImport`, являющийся частью интерфейсов импорта в LBS-платформу `Geo2Tag`. Загрузка данных в `MongoDB` осуществляется с помощью модуля `PyMongo` [64].

## **3.4. Тестирование решения**

Для проверки корректности работы решения, были разработаны тесты. Для тестирования использовался фреймворк `unittest`, также известный как `PyUnit` [65].

### **3.4.1. Юнит-тестирование**

Были протестированы следующие части плагина:

- Математические методы обработки данных:
  - Исходные данные: тестовые наборы числовых данных – случайно сгенерированный набор чисел;
  - Условие прохождения теста: соответствие вычисленного признака (среднее арифметическое, медиана) точному значению;
- Методы загрузки данных:
  - Тест-кейс 1 – возврат полученных данных:
    - Исходные данные: готовый для возврата объект;
    - Условие прохождения теста: соответствие возвращенного методом объекта известному;
  - Тест-кейс 2 – загрузка данных из удаленного источника:
    - Исходные данные: запрос на получение данных из удаленного источника, структура ответа (из документации удаленного источника);

- Условие прохождения теста: получение непустого ответа от удаленного источника, соответствие структуры ответа известной структуре;
- Методы формирования JSON-объектов необходимой структуры:
  - Исходные данные: необходимая структура JSON-объекта;
  - Условие прохождения теста: структура сформированного JSON-объекта соответствует необходимой.

### 3.4.2. Интеграционное тестирование

Интеграционное тестирование подразумевает тестирование функциональности объединенных модулей программного обеспечения, поэтому на данном этапе тестировалась следующая функциональность плагина:

- Последовательная загрузка и обработка данных, предоставление сформированных JSON-объектов на выходе:
  - Исходные данные: удаленный источник, необходимая структура JSON-объекта;
  - Условие прохождения теста: успешная загрузка данных, их обработка, возврат набора JSON-объектов, соответствующих заданной структуре;
- Запуск плагина через REST-интерфейс:
  - Исходные данные: REST-интерфейс платформы Geo2Tag, подключенный плагин;
  - Условие прохождения теста: выполнение POST-запроса к плагину, успешный ответ сервера;
- Добавление данных в базу LBS-платформы Geo2Tag:
  - Исходные данные: платформа Geo2Tag, подключенный плагин;
  - Условие прохождения теста: выполнение запроса к плагину через REST-интерфейс, ожидание его успешной работы. Выполнение запроса к базе данных – если есть новые данные – успешно.



### **3.5. Выводы**

В главе была подробно описана архитектура реализованного решения, алгоритм его работы. Представлены схемы JSON-объектов, получаемых из источников, а также JSON-объектов Point, хранящихся в базе данных платформы Geo2Tag.

Код решения размещен в открытом репозитории [66].

## **4. ИССЛЕДОВАНИЕ СВОЙСТВ РЕШЕНИЯ**

Для исследования принципиальной применимости и производительности разработанного плагина были поставлены эксперименты по измерению времени его работы при различных входных параметрах.

### **4.1. Условия эксперимента**

В качестве экспериментальной платформы использовался бытовой ноутбук со следующими характеристиками:

- Процессор Intel® Core™ i5-3210M с частотой 2.5 ГГц [67];
- Оперативное запоминающее устройство 4 Гб;
- Операционная система Ubuntu 16.04.2 LTS (Xenial Xerus) [68].

Так как работа плагина подразумевает загрузку данных, необходимо определить условия, связанные с использованным в экспериментах интернет-соединением:

- Заявленная скорость интернет-соединения – 100 Мбит/сек;
- Размер одного JSON-ответа на запрос к VK API – 0-40 Кбайт.

### **4.2. Используемые инструменты**

Чтобы запустить любой плагин импорта к платформе Geo2Tag необходимо выполнить POST запрос, передав необходимые для работы плагина аргументы. Аргументы, которые необходимо передать разработанному плагину:

- ChannelName – название канала, в который будут добавляться точки;
- Days\_to\_load – количество дней, образующих интервал от даты запуска плагина, за который будут загружены данные о записях, опубликованных в социальной сети «ВКонтакте»;
- Points\_to\_load – количество анализируемых точек Wi-Fi;

- Method – используемый метод анализа данных.

POST-запрос выполняется средствами служебной программы командной строки cURL [69].

Для автоматизации многократных запусков плагина, а также измерения времени его работы, были написаны два bash-сценария [70]:

- Run\_days – сценарий, принимающий на вход следующие аргументы:
  - Количество анализируемых точек Wi-Fi;
  - Минимальное количество дней для анализа;
  - Максимальное количество дней для анализа;
  - Используемый метод анализа.

Сценарий выполняет POST-запросы для запуска плагина с константным заданным числом точек Wi-Fi, используемым методом анализа данных и изменяющимся количеством дней для анализа.

- Run\_points – сценарий, принимающий на вход следующие аргументы:
  - Количество дней для анализа;
  - Минимальное количество анализируемых точек Wi-Fi;
  - Максимальное количество анализируемых точек Wi-Fi;
  - Используемый метод анализа.

Сценарий выполняет POST-запросы для запуска плагина с константным заданным количеством дней, используемым методом анализа данных и изменяющимся количеством анализируемых точек Wi-Fi.

Оба сценария транслируют ход работы в терминал. Данные о времени работы плагина, измеренные средствами системы с точностью до наносекунд, сохраняются локально в файл.

Время загрузки и обработки данных измеряется в модуле DataHandler с помощью модуля time [71] с точностью до наносекунд и сохраняется локально в файл.

#### **4.3. Результаты эксперимента**

Как было описано в разделе 3.3, работу плагина можно разделить на: получение данных, их обработку, добавление обработанных данных в базу. Так как интерфейсы импорта [35] выполняют загрузку данных в базу, после предоставления им данных для импорта, измерение времени работы плагина завершается. Поэтому время работы плагина состоит из:

- Времени получения данных с сайта открытых данных Санкт-Петербурга;
- Времени получения данных из социальной сети «ВКонтакте»;
- Времени обработки данных.

Так же плагин сохраняет данные о времени работы в файл локально, но время выполнения этих операций пренебрежимо мало, и не представляет интерес для исследования.

При проведении экспериментов использовались следующие интервалы значений параметров плагина:

- При тестировании зависимости времени работы от промежутка времени выборки: количество дней в выборке = 1 – 20, количество анализируемых точек = 89;
- При тестировании зависимости времени работы от количества анализируемых точек: количество дней в выборке = 20, количество анализируемых точек = 1 – 89 (3 – 89 при использовании метода кластеризации).

Линии тренда на графиках строились средствами Excel, используя метод наименьших квадратов. Экспериментальные данные, по которым строились графики, приведены в приложении Б.

Загрузка данных с сайта открытых данных Санкт-Петербурга выполняется при любых параметрах запуска плагина, причем всегда загружается тот же набор данных. Следовательно, ожидается отсутствие зависимости времени загрузки от промежутка времени выборки и от количества анализируемых точек Wi-Fi.

На рисунках 5, 6, 7, 8, 9, 10 представлены графики зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от промежутка времени выборки при методах обработки – среднее арифметическое (см. рис. 5), медиана (см. рис. 7), кластеризация (см. рис. 9), а также от количества анализируемых точек Wi-Fi при методах обработки – среднее арифметическое (см. рис. 6), медиана (см. рис. 8), кластеризация (см. рис. 10).



Рисунок 5 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от промежутка времени выборки при методе обработки – среднее арифметическое



Рисунок 6 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от количества анализируемых точек Wi-Fi при методе обработки – среднее арифметическое



Рисунок 7 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от промежутка времени выборки при методе обработки – медиана



Рисунок 8 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от количества анализируемых точек Wi-Fi при методе обработки – медиана

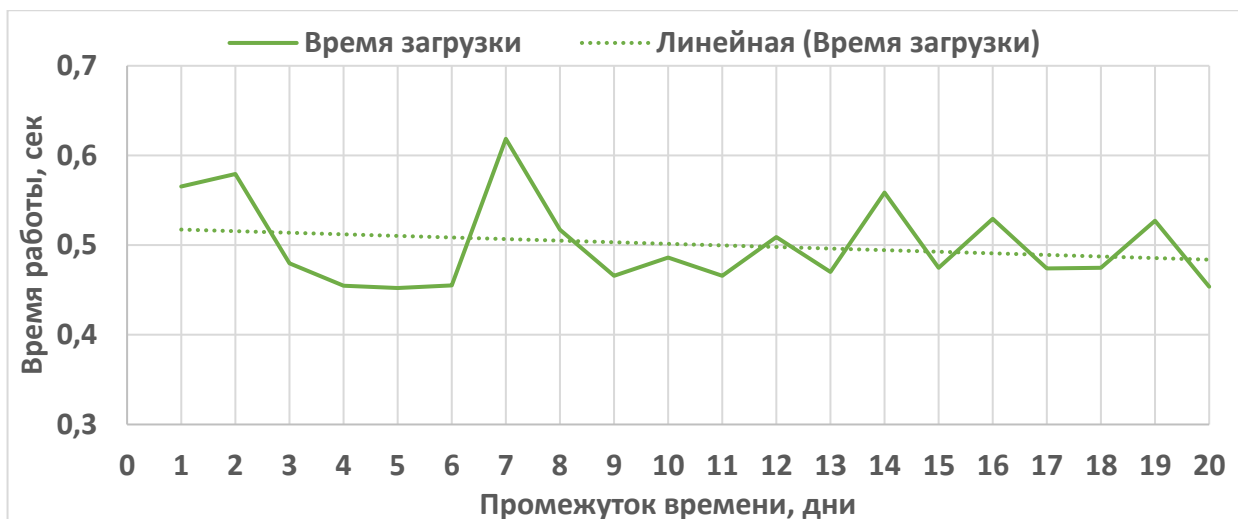


Рисунок 9 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от промежутка времени выборки при методе обработки – кластеризация



Рисунок 10 – график зависимости времени загрузки данных с сайта открытых данных Санкт-Петербурга от количества анализируемых точек Wi-Fi при методе обработки – кластеризация

Как видно на рисунках 5, 6, 7, 8, 9, 10 время загрузки данных с сайта открытых данных Санкт-Петербурга, ожидаемо, не изменяется. Отклонения на графике вызваны нестабильностью интернет соединения.

#### 4.3.1. Зависимость времени работы плагина от промежутка времени выборки

В данном подразделе проанализирована зависимость времени работы плагина от промежутка времени выборки, основываясь на полученных экспериментальных данных.

#### 4.3.1.1. Метод обработки: среднее арифметическое

На рис. 11 изображён график зависимости времени работы плагина от промежутка времени выборки, за который запрашиваются записи, опубликованные в социальной сети «ВКонтакте», при выбранном методе обработки данных – среднее арифметическое.

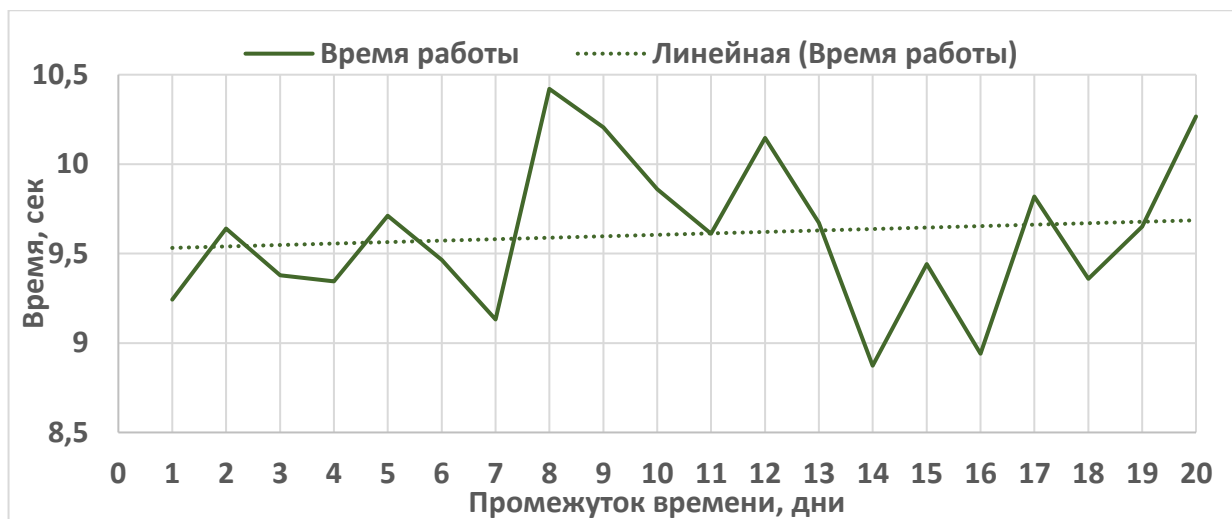


Рисунок 11 – график зависимости времени работы плагина от промежутка времени выборки

Как видно на графике, изображенном на рис. 11, время работы плагина не зависит от промежутка времени выборки. Или зависимость линейная, но коэффициент при переменной близок к 1. Для того, чтобы точно определить зависимость, были исследованы зависимости составляющих общей времени работы плагина от промежутка времени выборки: на рис. 12 представлена зависимость времени загрузки данных из социальной сети «ВКонтакте», на рис. 13 представлена зависимость времени обработки данных.



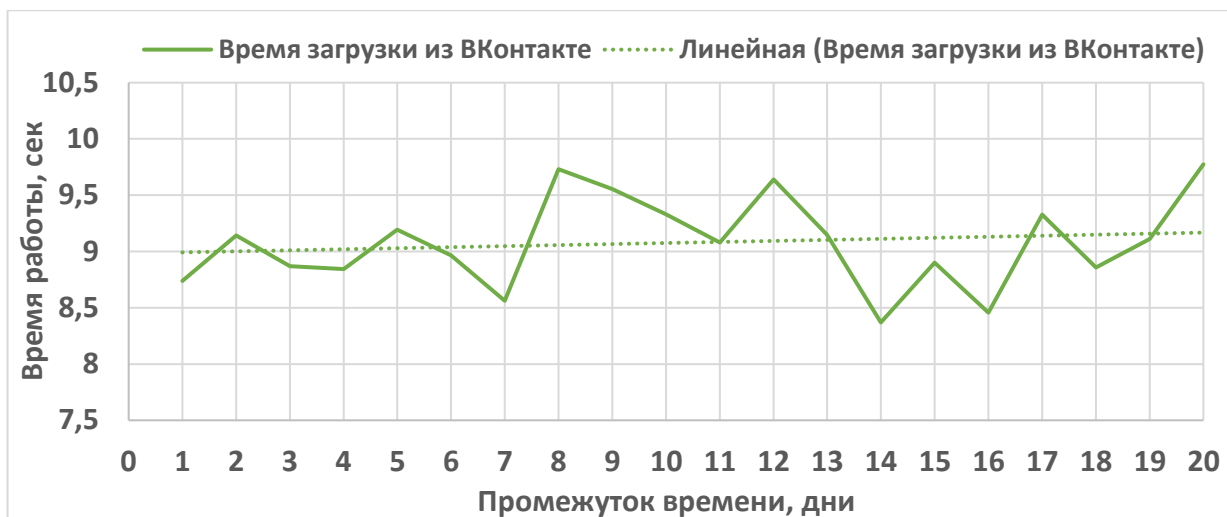


Рисунок 12 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки

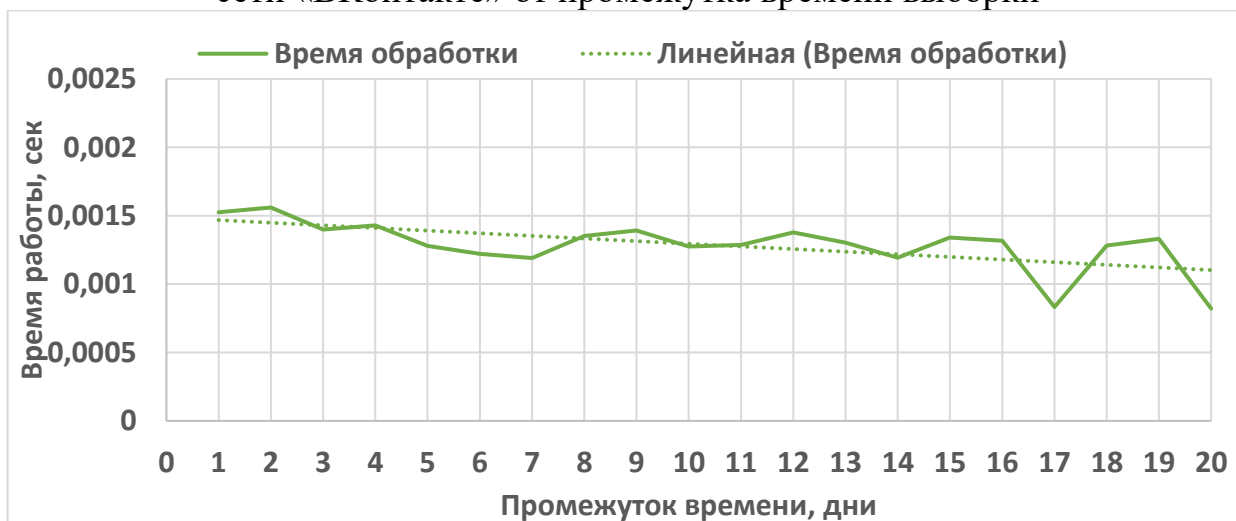


Рисунок 13 – график зависимости времени обработки данных от промежутка времени выборки

Как видно на рис. 12, зависимость времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки похожа на зависимость времени работы плагина. Также стоит отметить, что время загрузки данных сравнимо с общим временем работы плагина.

Как видно на рис. 13, зависимость времени обработки данных от промежутка времени выборки не наблюдается. Это логично, если учесть тот факт, что количество точек неизменно. Также видно, что время обработки пренебрежимо мало, по сравнению с временем загрузки данных – оно меньше на четыре порядка.

Так как количество постов, опубликованных в социальной сети «ВКонтакте» растет со временем, логично ожидать увеличение времени загрузки данных о них, но, как было указано в 4.1 – максимальный размер ответа на данный запрос, полученный во время эксперимента, составил 40 Кбайт, соответственно рост времени загрузки данных такого размера при заданных условиях эксперимента (заявленная скорость интернет соединения 100 Мбит/сек) сложно заметить.

Из графиков, изображенных на рисунках 11 - 13 можно сделать вывод, что подавляющую часть времени работы плагина составляет время загрузки данных из социальной сети «ВКонтакте». Время работы плагина не зависит от промежутка времени выборки.

#### **4.3.1.2. Метод обработки: медиана**

Основываясь на результатах исследования результатов эксперимента, описанных в подразделе 4.3.1.1, логично ожидать отсутствие зависимости времени работы плагина от промежутка времени выборки при использовании метода обработки «медиана», так как время обработки данных пренебрежимо мало, а подавляющую часть времени работы плагина занимает загрузка данных из социальной сети «ВКонтакте».

На рис. 14 представлен график зависимости времени работы плагина от промежутка времени выборки, на рис. 15 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки, на рис. 16 – график зависимости времени обработки данных от промежутка времени выборки.

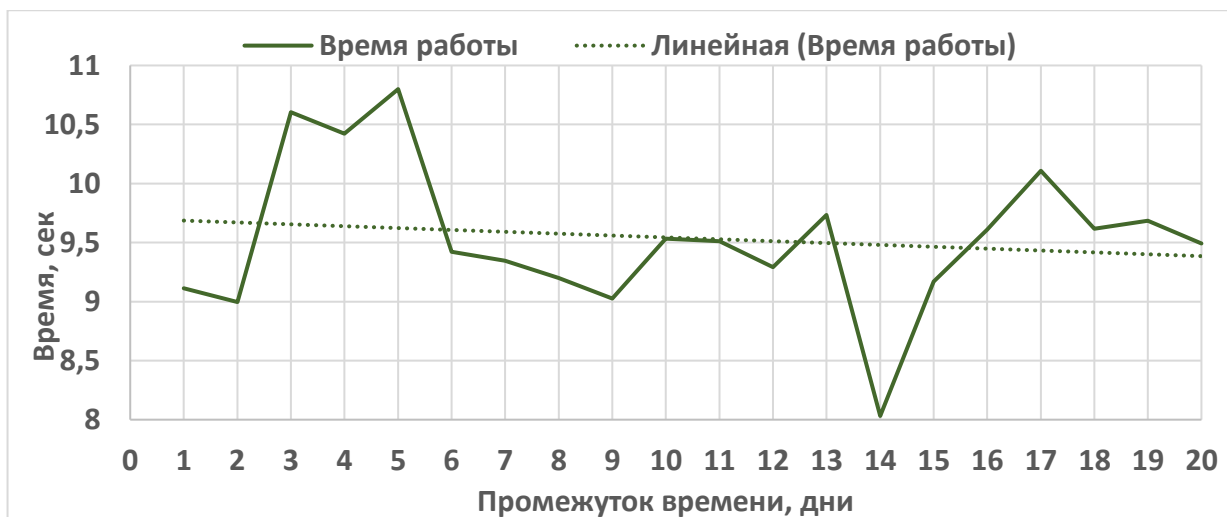


Рисунок 14 – график зависимости времени работы плагина от промежутка времени выборки

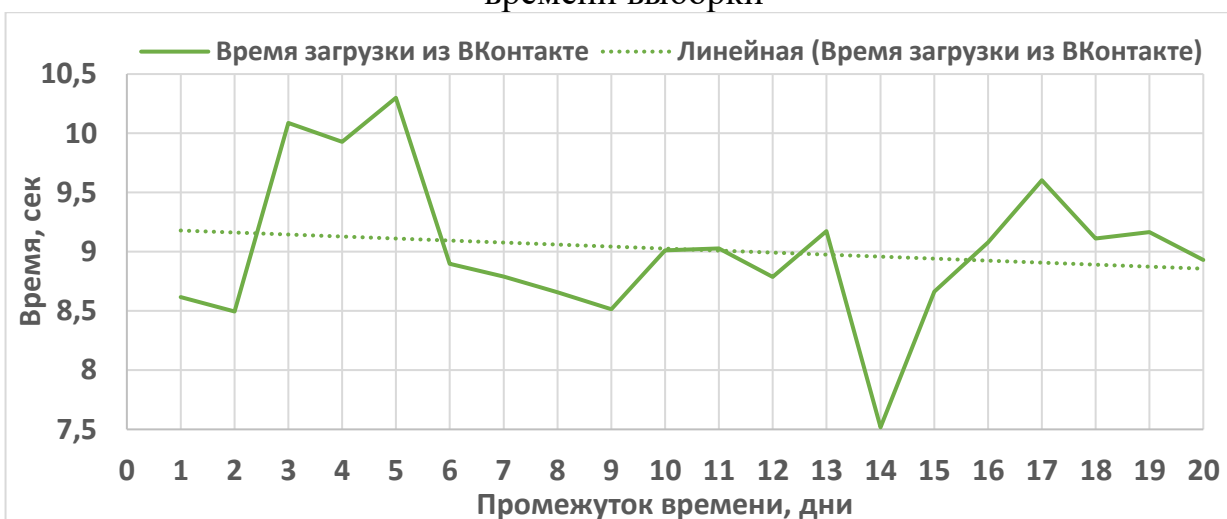


Рисунок 15 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки



Рисунок 16 – график зависимости времени обработки данных от промежутка времени выборки

Результаты, представленные на рисунках 14, 15 и 16, аналогичны результатам, полученным в подразделе 4.3.1.1, то есть:

- Время обработки данных пренебрежимо мало, по сравнению с временем загрузки данных. Зависимости от промежутка времени выборки нет (так как количество анализируемых точек Wi-Fi неизменно);
- Время загрузки данных из социальной сети «ВКонтакте» можно считать не зависящим от промежутка времени выборки, в связи с малым объемом загружаемых данных;
- Время работы плагина не зависит от промежутка времени выборки.

#### **4.3.1.3. Метод обработки: кластеризация**

Основываясь на результатах исследования результатов эксперимента, описанных в подразделах 4.3.1.1 и 4.3.1.2, логично ожидать отсутствие зависимости времени работы плагина от промежутка времени выборки при использовании метода обработки «кластеризация», так как время обработки данных пренебрежимо мало, а подавляющую часть времени работы плагина занимает загрузка данных из социальной сети «ВКонтакте».

На рис. 17 представлен график зависимости времени работы плагина от промежутка времени выборки, на рис. 18 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки, на рис. 19 – график зависимости времени обработки данных от промежутка времени выборки.

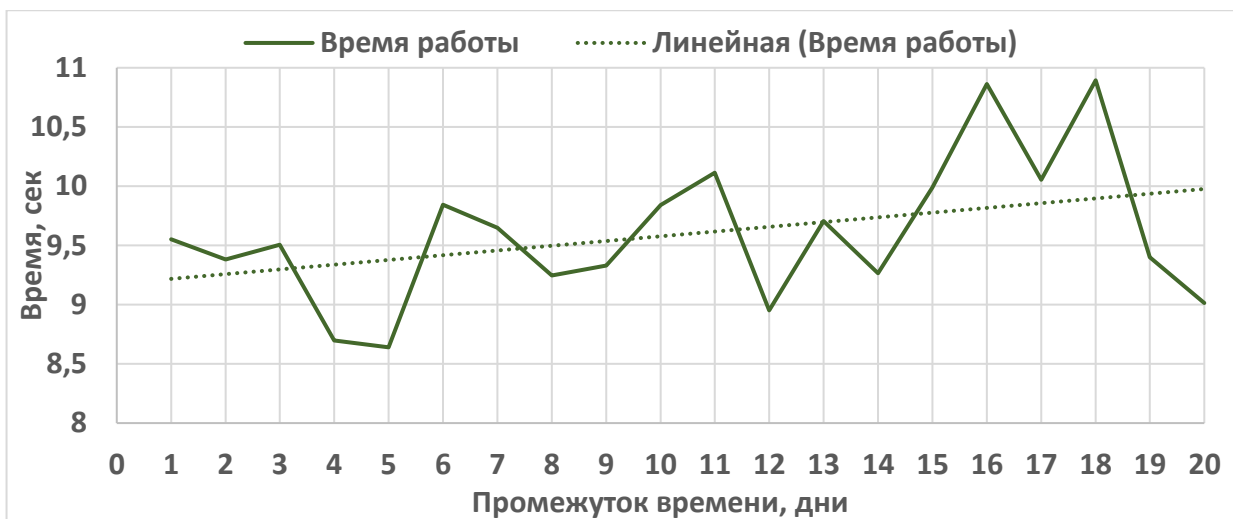


Рисунок 17 – график зависимости времени работы плагина от промежутка времени выборки



Рисунок 18 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от промежутка времени выборки



Рисунок 19 – график зависимости времени обработки данных от промежутка времени выборки

Результаты, представленные на рисунках 17, 18 и 19, аналогичны результатам, полученным в подразделах 4.3.1.1 и 4.3.1.3, то есть:

- Время обработки данных пренебрежимо мало, по сравнению с временем загрузки данных. Зависимости от промежутка времени выборки нет (так как количество анализируемых точек Wi-Fi неизменно);
- Время загрузки данных из социальной сети «ВКонтакте» можно считать не зависящим от промежутка времени выборки, в связи с малым объемом загружаемых данных;
- Время работы плагина не зависит от промежутка времени выборки.

В результате экспериментов было установлено, что при любом методе обработки данных, время работы плагина не зависит от промежутка времени выборки.

#### **4.3.2. Зависимость времени работы плагина от количества анализируемых точек Wi-Fi**

В данном подразделе проанализирована зависимость времени работы плагина от количества анализируемых точек Wi-Fi, основываясь на полученных экспериментальных данных.

В разделе 4.3.1. выяснилось, что подавляющую часть времени работы плагина составляет загрузка данных из социальной сети «ВКонтакте», а так как количество запросов к социальной сети равно количеству анализируемых точек Wi-Fi, следует ожидать линейную зависимость времени работы плагина от количества точек.

##### **4.3.2.1. Метод обработки: среднее арифметическое**

На рис. 20 изображён график зависимости времени работы плагина от количества анализируемых точек Wi-Fi, при выбранном методе обработки данных – среднее арифметическое.



Рисунок 20 – график зависимости времени работы плагина от количества анализируемых точек Wi-Fi

Как видно на графике, изображенном на рис. 20, время работы плагина линейно зависит от количества анализируемых точек Wi-Fi. С помощью средств Excel, было получено уравнение линии тренда:

$$y = 0.1036x + 0.5379 \quad (1)$$

На рис. 21 представлена зависимость времени загрузки данных из социальной сети «ВКонтакте», на рис. 22 представлена зависимость времени обработки данных.



Рисунок 21 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от количества анализируемых точек Wi-Fi



Рисунок 22 – график зависимости времени обработки данных от количества анализируемых точек Wi-Fi

Как видно на рис. 21, время загрузки данных из социальной сети «ВКонтакте», также линейно зависит от количества анализируемых точек Wi-Fi, что ожидалось, ведь именно загрузка этих данных занимает подавляющее большинство времени работы плагина.

Как видно на рис. 22, зависимость времени обработки данных с выделением среднего арифметического набора от количества анализируемых точек Wi-Fi линейная. Уравнение линии тренда:

$$y = 0.00001595x + 0.00001054 \quad (2)$$

Также видно, что время обработки пренебрежимо мало, по сравнению с временем загрузки данных – оно меньше на четыре порядка.

#### 4.3.2.2. Метод обработки: медиана

На рис. 23 изображён график зависимости времени работы плагина от количества анализируемых точек Wi-Fi, при выбранном методе обработки данных – медиана.





Рисунок 23 – график зависимости времени работы плагина от количества анализируемых точек Wi-Fi

Как видно на графике, изображенном на рис. 23, время работы плагина линейно зависит от количества анализируемых точек Wi-Fi. С помощью средств Excel, было получено уравнение линии тренда:

$$y = 0.0965x + 0.6044 \quad (3)$$

На рис. 24 представлена зависимость времени загрузки данных из социальной сети «ВКонтакте», на рис. 25 представлена зависимость времени обработки данных.



Рисунок 24 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от количества анализируемых точек Wi-Fi



Рисунок 25 – график зависимости времени обработки данных от количества анализируемых точек Wi-Fi

Как видно на рис. 24, время загрузки данных из социальной сети «ВКонтакте», также линейно зависит от количества анализируемых точек Wi-Fi, что ожидалось, ведь именно загрузка этих данных занимает подавляющее большинство времени работы плагина.

Как видно на рис. 25, зависимость времени обработки данных с выделением медианы набора от количества анализируемых точек Wi-Fi линейная. Уравнение линии тренда:

$$y = 0.00001474x + 0.00005427 \quad (4)$$

Время обработки, по сравнению с временем загрузки данных, также пренебрежимо мало – оно меньше на четыре порядка.

#### 4.3.2.3. Метод обработки: кластеризация

На рис. 26 изображён график зависимости времени работы плагина от количества анализируемых точек Wi-Fi, при выбранном методе обработки данных – кластеризация.



Рисунок 26 – график зависимости времени работы плагина от количества анализируемых точек Wi-Fi

Как видно на графике, изображенном на рис. 26, время работы плагина линейно зависит от количества анализируемых точек Wi-Fi. С помощью средств Excel, было получено уравнение линии тренда:

$$y = 0.096x + 0.5306 \quad (5)$$

На рис. 27 представлена зависимость времени загрузки данных из социальной сети «ВКонтакте», на рис. 28 представлена зависимость времени обработки данных.

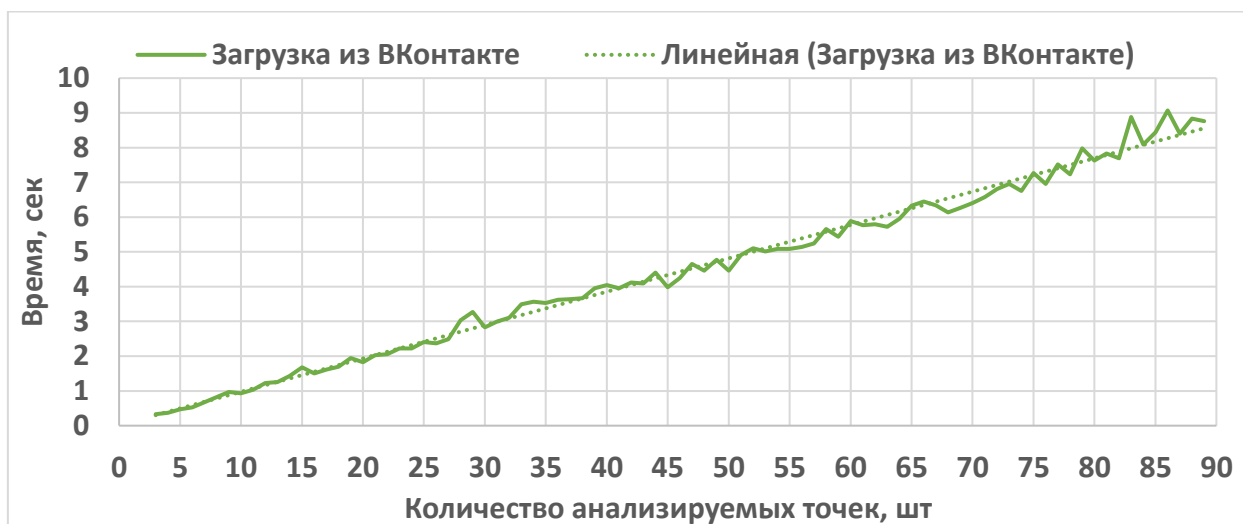


Рисунок 27 – график зависимости времени загрузки данных из социальной сети «ВКонтакте» от количества анализируемых точек Wi-Fi



Рисунок 28 – график зависимости времени обработки данных от количества анализируемых точек Wi-Fi

Как видно на рис. 27, время загрузки данных из социальной сети «ВКонтакте», также линейно зависит от количества анализируемых точек Wi-Fi, что ожидалось, ведь именно загрузка этих данных занимает подавляющее большинство времени работы плагина.

Как видно на рис. 28, зависимость времени обработки данных методом кластеризации от количества анализируемых точек Wi-Fi степенная. Это соответствует теоретическому описанию сложности кластеризации методом k-средних [72]. Уравнение линии тренда:

$$y = 0.0000606x^{1.2543} \quad (6)$$

Время обработки, по сравнению с временем загрузки данных, также пренебрежимо мало – оно меньше примерно в 500 раз.

#### 4.3.2.4. Сравнение методов обработки данных

На рис. 29 изображен график зависимости времени работы плагина от количества анализируемых точек при каждом методе обработки данных.

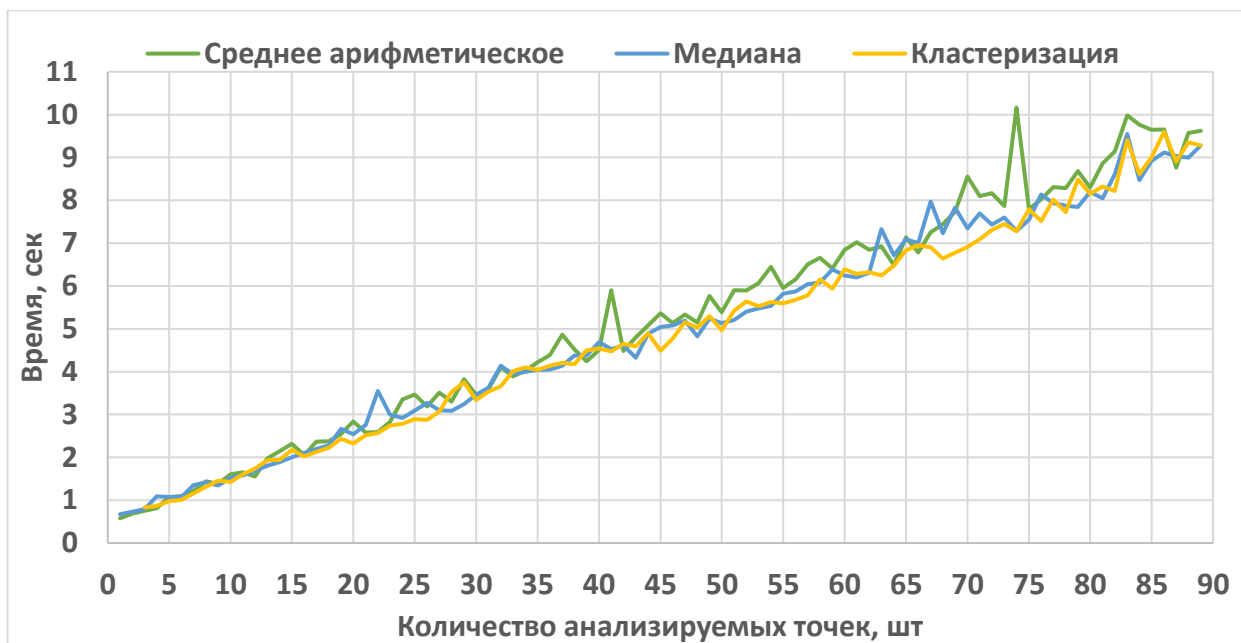


Рисунок 29 – график зависимости времени работы плагина от количества анализируемых точек при различных методах обработки данных

Как видно на рис. 29, время работы плагина при разных методах обработки данных почти не отличается, и это ожидаемо, т.к. выше выяснилось, что подавляющую часть времени работы плагина занимает загрузка данных из социальной сети «ВКонтакте».

Поэтому стоит сравнить время обработки данных разными методами – график изображен на рис. 30.

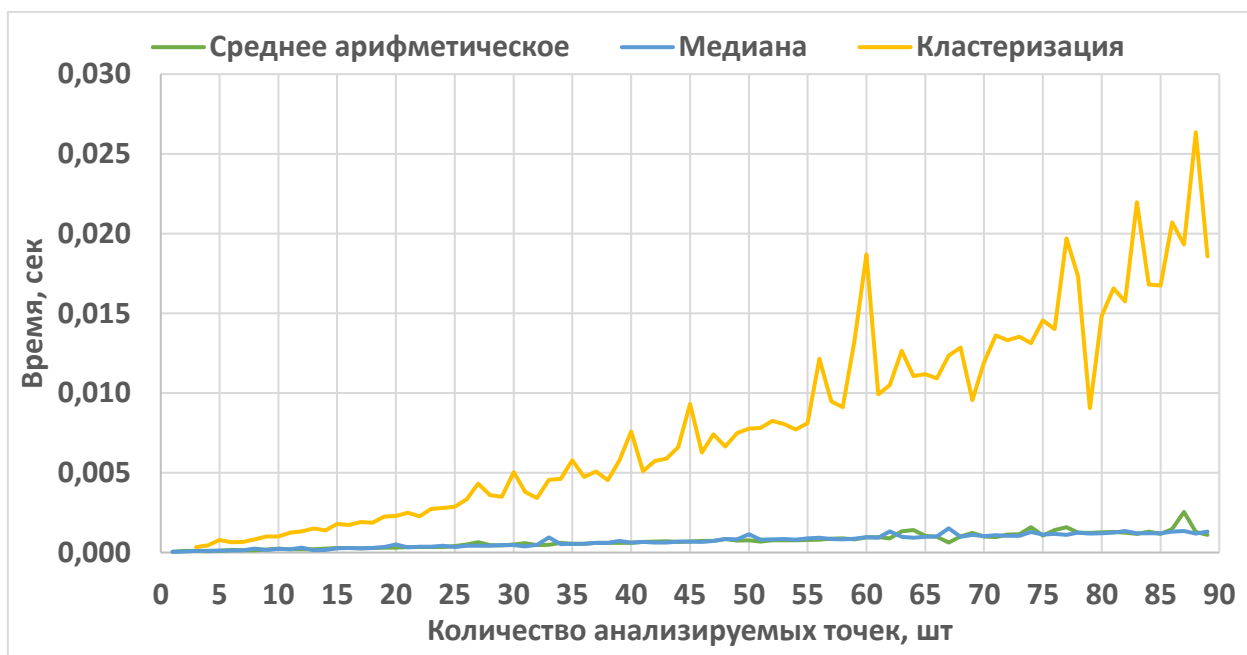


Рисунок 30 – график зависимости времени обработки данных различными методами от количества анализируемых точек

Как видно на рис. 30, время обработки данных методом кластеризации больше, чем время обработки данных с выделением среднего арифметического или медианы набора. В наибольших значениях количества анализируемых точек, использовавшихся в эксперименте, разница во времени обработки данных методом кластеризации и методами «среднее арифметическое» и «медиана» достигает 15-20 раз.

#### **4.4. Выводы**

В ходе экспериментов было установлено, что время работы плагина не зависит от промежутка времени выборки, так как подавляющую часть времени работы плагина занимает получение данных из социальной сети «ВКонтакте», в связи с чем наблюдается линейная зависимость между временем работы плагина и количеством анализируемых точек Wi-Fi.

Было проведено сравнение реализованных методов обработки данных. Метод кластеризации, в виду его сложности, работает дольше, чем выделение дополнительного признака при использовании методов вычисления среднего арифметического и медианы.

## **5. БЕЗОПАСНОСТЬ ЖИЗНЕДЕЯТЕЛЬНОСТИ**

Эргономика — наука о приспособлении должностных обязанностей, рабочих мест, предметов и объектов труда, а также компьютерных программ для наиболее безопасного и эффективного труда работника, исходя из физических и психических особенностей человеческого организма.

Важным аспектом при разработке программ является разработка пользовательского интерфейса, ведь он для подавляющего большинства пользователей и является программой. Удобство работы с программой напрямую зависит от пользовательского интерфейса, от прилагаемых усилий для его продуманной разработки.

Но не стоит забывать и о другой стороне работы пользователя с программой — диалога человека и системы как такового. Продуманность диалога разработчиками как ничто другое влияет на удобство выполнения пользователем поставленных перед ним задач.

Далее рассмотрены семь принципов диалога, применяемых при проектировании и оценке программ и интерактивных систем, которые также могут рассматриваться как цели при проектировании и критерии оценки диалога. Стоит отметить, что эти принципы не являются независимыми и могут семантически пересекаться. Применимость и приоритет каждого принципа зависят от области применения, групп пользователей и других факторов.

### **5.1. Приемлемость организации диалога для выполнения поставленных задач**

Разработанная система соответствует поставленным перед ней задачам, если она помогает пользователю выполнять их, то есть функциональность и диалог основаны на особенностях задач, а не на выборе методов их выполнения.

Полезно в процессе диалога предоставлять статус выполнения задач, в том числе и успешное их завершение. Например, если сроки выполнения задачи ограничены, то необходимо отображать оставшееся время.

Но в то же время необходимо избегать предоставления пользователю информации, не нужной для выполнения задачи. Например, если путешественник хочет заказать гостиничный номер на конкретную дату, то необходимо отражать только гостиниц со свободными номерами на указанный день. Информация о гостиницах, не имеющих свободные номера в назначенную дату является бесполезной, и может лишь привести к снижению производительности пользователя, добавить ему лишнюю умственную нагрузку.

Формат ввода-вывода должен соответствовать его цели. Например, если пользователь должен выбрать дату, то необходимо предоставить ему инструмент, позволяющий ввести только её – то есть календарь. Если система производит конвертацию валют, необходимо выводить данные с точностью до двух знаков после запятой.

Если для выполнения задачи необходимы настройки системы, то такие настройки должны предусматривать значения по умолчанию. Например, в случае покупки железнодорожного билета, пользователи чаще всего покупают билет от станции отправления, значит есть смысл сделать эту станцию начальным пунктом по умолчанию.

Необходимо сократить диалог до минимума, необходимого для выполнения поставленной задачи. Например, при заполнении данных о контакте, при вводе почтового индекса автоматически должен указываться город.



## **5.2. Информативность**

Диалог должен быть информативен в такой степени, чтобы пользователю было ясно, в каком диалоге он находится, какие действия и как могут быть выполнены.

Информация, представленная пользователю на любом шаге диалога, должна способствовать его завершению. Например, система бронирования мест в гостинице должны позволять пользователю вводить необходимые данные и использовать кнопки «Далее» и «Назад» для осуществления необходимых шагов диалога.

При работе с программой необходимость обращения к документации должна быть сведена к минимуму. Например, программа должна предоставлять меню, пункты которого ясно отражают типичные задачи пользователей, решение которых поддерживается программой.

Пользователя необходимо держать в курсе возможных изменений в состоянии интерактивной системы с помощью сообщений об ожидании ввода, краткого обзора предстоящих этапов диалога. Например, при покупке в интернет-магазине необходимо отобразить пользователю все необходимые этапы, которые должны быть выполнены для покупки продукта.

Диалог должен быть разработан так, чтобы взаимодействие с программой было всегда понятно пользователю. Например, программа, проигрывающая аудио-видео файлы, должна предоставлять пользователю управление с помощью значков, изображающих общепринятые отображения кнопок «воспроизведение», «остановка», «пауза», «перемотка вперед» и так далее.

Интерактивная система должна предоставлять пользователю информацию об используемых форматах и единицах измерения. К примеру, банковская система обязательно должна отображать валюту.

### **5.3. Соответствие ожиданиям пользователя**

Диалог соответствует ожиданиям пользователей, если он соответствует предсказуемым, зависящим от области применения требованиям пользователя.

В программе должна быть использована терминология, свойственная области применения программы. Например, система, использующая банкирами, должна использовать банковскую терминологию.

Пользователь должен быть обеспечен оперативной и удобной связью, соответствующей его ожиданиям. Например, при установке программы, пользователь должен быть уведомлен об успешном завершении инсталляции.

Если реальное время реакции системы на действия пользователя значительно отклоняется от времени, ожидаемого пользователем, то пользователь должен быть проинформирован об этом. Например, если пользователь ищет интересующие его воздушные рейсы на вебсайте авиакомпании, и нажимает кнопку поиска, то можно проинформировать пользователя о том, что поиск ведется, но в связи с загруженностью сервера результаты поиска будут отображены на экране лишь по истечении 30 секунд.

Форматы должны соответствовать культурным и лингвистическим соглашениям. Например, программа должна использовать левое упорядочение для английского языка и правое упорядочение для арабских языков.

### **5.4. Пригодность для обучения**

Диалог является пригодным для обучения, если он помогает пользователю изучать инструкции по использованию системы.

Если из-за нечастого использования диалога или характеристики пользователя необходимо повторное изучение диалога, то в таком случае должна быть обеспечена соответствующая поддержка. Например, программа для бухгалтерского учета включает справочную систему, которая посредством диалога дает пользователю поэтапные указания по созданию годового баланса.

Пользователю должны быть оказана соответствующая поддержка при его ознакомлении с диалогом. Например, при первом использовании некоторого функционала, программа должны выдавать справку по его использованию.

Обратная связь и пояснения должны помочь пользователю в формировании понимания системы. Например, программа сканирования должны подсказывать все действия, требуемые для сканирования печатного документа и конвертирования его в электронный текстовый документ, а также показывать текущее действие.

### **5.5. Контролируемость**

Диалог является контролируемым, если пользователь имеет возможность инициировать и контролировать направления и темп диалогового взаимодействия до того момента, пока цель не достигнута.

Темп взаимодействия между пользователем и системой не должен зависеть от функциональных возможностей и ограничений системы. Он должен определяться пользователем и находиться под его контролем в соответствии с потребностями пользователя. Например, почтовый клиент позволяет пользователю отправлять сообщения, сохраняет частично завершенные сообщения видимыми и доступными для редактирования до тех пор, пока пользователь не решит, отправить или удалить сообщение, независимо от фактического времени, затраченного пользователем на завершение отправки сообщения.

Пользователь должен иметь возможность выбора вариантов продолжения диалога. Например, банковское приложение, при выполнении пользователем платежа, должно давать возможность сохранения шаблона платежа для дальнейшего использования.

Если операция является обратимой, то пользователь должен иметь возможность отменить, по крайней мере, последний этап диалога. Например, приложение-редактор должно позволять пользователю отменить последнюю произведенную операцию редактирования.

### **5.6. Устойчивость к ошибкам**

Диалог является устойчивым к ошибкам, если, несмотря на очевидные ошибки на входе, предполагаемый результат может быть достигнут или без корректирующих воздействий, или с минимальными корректирующими воздействиями пользователя. Устойчивость к ошибкам может быть достигнута посредством контроля и исправления ошибок, а также посредством управления обработкой ошибок для их исправления.

Система должна помогать пользователю в обнаружении и предупреждении ошибок на входе. Например, программа должна указывать пользователю на обязательные поля, которые не были им заполнены.

Система должна предупреждать любые действия пользователя, которые могут привести к возникновению неопределенных состояний системы или её отказу. Например, диалог при печати документа объемом 20 страниц должен позволять пользователю распечатать страницы с 1 по 20.

Если ошибка произошла, то её происхождение должно быть объяснено пользователю для облегчения её исправления. Например, DVD-проигрыватель в случае ошибки предоставляет сообщение «Вы нажали кнопку «Play», но диск для проигрывания в дисководе отсутствует».

Пользователю нужно предоставить возможность отложить исправление ошибки, если её исправление не обязательно для продолжения диалога. Например, при заполнении формы для регистрации на сайте, при вводе не валидного адреса электронной почты, пользователь может продолжать вводить свои данные в другие поля.

### **5.7. Адаптируемость к индивидуальным особенностям пользователя**

Диалог является адаптируемым к индивидуальным особенностям применения, если пользователи могут внести изменения в формат взаимодействия с системой и в формы представления информации для того, чтобы удовлетворить свои индивидуальные возможности и потребности.

Система должна предоставлять пользователю возможность выбора альтернативных форм представления информации, если эта система адаптирована к индивидуальным потребностям различных пользователей, например, слабовидящие пользователи должны иметь возможность передвигаться по экрану, используя программное средство, конвертирующее информацию на экране в акустический сигнал.

Степень подробности объяснений должна зависеть от индивидуального уровня знаний пользователя. Например, приложение должно позволять включать опцию подсказок, инициируемую системой для малоопытных пользователей.

Пользователь должен иметь возможность выбрать уровни и способы взаимодействия, которые лучше всего соответствуют его потребностям. Например, текстовые редакторы поддерживают функцию сохранения документа посредством опции меню, значка или клавишной комбинации.

### **5.8. Выводы**

Продуманность диалогов взаимодействия системы с пользователем напрямую влияет на удобство работы с системой, легкость выполнения поставленных перед пользователем задач. Правильная реакция системы на действия пользователя, а также своевременная помощь ему в работе – залог успеха программы.

## ЗАКЛЮЧЕНИЕ

По итогам работы были получены следующие результаты:

- Был проведен обзор предметной области, по результатам которого были даны определения и основная информация об открытых данных, контекстно-зависимых вычислениях, геоконтекстной разметке. Также был проведен обзор существующих решений по обработке данных и работе с геоданными – Big Data решения и ГИС, по результатам которого было показано, что они не подходят для решения поставленной задачи;
- Был проведен обзор и сравнительный анализ LBS-платформ, по результатам которого была показана целесообразность использования Geo2Tag;
- Была спроектирована архитектура плагина к LBS-платформе Geo2Tag;
- На основе спроектированной архитектуры был реализован плагин, решающий следующие подзадачи:
  - Получение данных из двух источников – сайта открытых данных Санкт-Петербурга и социальной сети «ВКонтакте»;
  - Обработка полученных данных, с выделением дополнительного признака, основываясь на числовом показателе;
  - Добавление обработанных данных в базу LBS-платформы Geo2Tag;
- Проведено тестирование решения;
- Приведены и проанализированы результаты исследования полученного решения: время работы плагина не зависит от промежутка времени выборки, но линейно зависит от количества анализируемых точек Wi-Fi.

Поставленные задачи были решены, цель работы была достигнута.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Marr B. Big Data: 20 Mind-Boggling Facts Everyone Must Read //Forbes Magazine. – 2015.
2. Dey A. K. Understanding and using context //Personal and ubiquitous computing. – 2001. – Т. 5. – №. 1. – С. 4-7.
3. Basiri A. et al. Challenges of location-based services market analysis: current market description //Progress in Location-Based Services 2014. – Springer International Publishing, 2015. – С. 273-282.
4. Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data". The Semantic Web. Lecture Notes in Computer Science. 4825. p. 722. doi:10.1007/978-3-540-76298-0\_52. ISBN 978-3-540-76297-3.
5. Официальный сайт OpenSource // URL: <https://opensource.org/> (дата обращения: 01.05.2017).
6. Suber P. Open access overview // URL: <http://www.earlham.edu/~peters/fos/overview.htm>. (дата обращения: 01.05.2017)
7. Официальный сайт открытых данных США // URL: <https://www.data.gov/> (дата обращения: 01.05.2017).
8. Официальный сайт открытых данных Великобритании // URL: <https://data.gov.uk/> (дата обращения: 01.05.2017).
9. Bray T. The javascript object notation (json) data interchange format. – 2014.
10. Bray T. et al. Extensible Markup Language (XML) 1.0 5th Edition, W3C recommendation, November 2008.
11. The World Wide Web Consortium (W3C) // URL: <https://www.w3.org/> (дата обращения: 01.05.2017).
12. World Wide Web Consortium et al. RDF 1.1 Primer. – 2014.
13. Adida B., Commons C., Birbeck M. RDFa Core 1.1. – 2007.
14. Morris W. et al. American heritage dictionary of the English language (5<sup>th</sup> ed). – American heritage, 2011.

15. Microsoft Excel // URL: <https://products.office.com/ru-ru/excel> (дата обращения: 01.05.2017).
16. LibreOffice Calc // URL: <http://www.libreoffice.org/discover/calc/> (дата обращения: 01.05.2017).
17. Shafranovich Y. Common format and MIME type for comma-separated values (CSV) files. – 2005.
18. [MS-DOC]: Word (.doc) Binary File Format // URL: [https://msdn.microsoft.com/en-us/library/office/cc313153\(v=office.12\).aspx](https://msdn.microsoft.com/en-us/library/office/cc313153(v=office.12).aspx) (дата обращения: 01.05.2017).
19. Weir R. OpenDocument format: The standard for office documents //IEEE Internet Computing. – 2009. – Т. 13. – №. 2. – С. 83-87.
20. Rice F. Introducing the office (2007) open XML file formats //Microsoft Developer Network. – 2006.
21. Iso T. C. 171/SC 2: ISO 32000–1: 2008 Document Management-Portable Document Format-Part 1: PDF 1.7.
22. Berjon R. et al. HTML 5.1 specification //W3C Working Draft. – 2014.
23. Schilit BN, Theimer MM. Disseminating active map information to mobile hosts (англ.). — IEEE Network, 1994
24. Schmidt A, Beigl M, Gellersen HW. There is more to context than location (англ.). — Computers & Graphics Journal, Elsevier, 1999.
25. Schmidt A, van Laerhoven K. How to build smart appliances? (англ.). — IEEE Personal Communications, 2001.
26. Zaslavskiy M., Mouromtsev D. Implementation of the new REST API for open source LBS-platform Geo2Tag //Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), 2015. – IEEE, 2015. – С. 125-130.
27. E.S. Lohan, T. Kauppinen, and S.B. Chandra Debnath, “A Survey of People Movement Analytics Studies in the Context of Smart Cities”, in Proc. of FRUCT 2016 conference, Jyväskylä, Finland. Nov 2016



28. Google intro to geocoding // URL: <https://developers.google.com/maps/documentation/geocoding/intro> (дата обращения: 01.05.2017).

29. Quercia, Daniele; Lathia, Neal; Calabrese, Francesco; Di Lorenzo, Giusy; Crowcroft, Jon (2010). Recommending Social Events from Mobile Phone Location Data (PDF). 2010 IEEE International Conference on Data Mining. p. 971. doi:10.1109/ICDM.2010.152. ISBN 978-1-4244-9131-5.

30. Anagog LBS-platform website // URL: <http://anagog.com> (дата обращения: 01.05.2017).

31. Flask framework // URL: <http://flask.pocoo.org/> (дата обращения: 01.05.2017).

32. MongoDB website // URL: <https://www.mongodb.com/> (дата обращения: 01.05.2017).

33. Apache http server project // URL: <https://httpd.apache.org/> (дата обращения: 01.05.2017).

34. Introducing R. Representational State Transfer (REST).

35. Заславский М.М., Баландин С.И. Метод импорта и обработки открытых данных в LBS-платформе // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 5. С. 816–822. doi: 10.17586/2226-1494-2016-16-5-816-822

36. Boyd D., Crawford K. Six provocations for big data // A decade in internet time: Symposium on the dynamics of the internet and society. – Oxford : Oxford Internet Institute, 2011. – Т. 21.

37. Chen M. et al. Big data: related technologies, challenges and future prospects. – Heidelberg : Springer, 2014. – С. 2-9.

38. Apache Hadoop // URL: <http://hadoop.apache.org/> (дата обращения: 01.05.2017).

39. Apache Spark // URL: <https://spark.apache.org/> (дата обращения: 01.05.2017).

40. Apache Storm // URL: <http://storm.apache.org/> (дата обращения: 01.05.2017).
41. Teradata Big Data // URL: <http://bigdata.teradata.com/> (дата обращения: 01.05.2017).
42. Apache Cassandra // URL: <http://cassandra.apache.org/> (дата обращения: 01.05.2017).
43. Dijcks J. P. Oracle: Big data for the enterprise //Oracle White Paper. – 2012.
44. First G. I. S., Second G. I. S., Third G. I. S. Geographic information systems as an integrating technology: context, concepts, and definitions. – 1996.
45. ArcGIS website // URL: <https://www.arcgis.com/features/index.html> (дата обращения: 01.05.2017).
46. GIS as a service // URL: <https://learn.thundermaps.com/blog-posts/what-is-gis-as-a-service-gaas/> (дата обращения: 01.05.2017)
47. Google Maps // URL: <https://www.google.ru/maps> (дата обращения: 01.05.2017).
48. Apple Maps // URL: <https://www.apple.com/ios/maps/> (дата обращения: 01.05.2017).
49. Bing Maps // URL: <https://www.bing.com/maps> (дата обращения: 01.05.2017).
50. OpenStreetMap // URL: <https://www.openstreetmap.org/> (дата обращения: 01.05.2017).
51. CartoDB // URL: <https://carto.com/> (дата обращения: 01.05.2017).
52. MapBox // URL: <https://www.mapbox.com/> (дата обращения: 01.05.2017).
53. Azoft LBS-platform // URL: <http://www.azoft.com/solutions/lbs/> (дата обращения: 01.05.2017).
54. TomTom LBS-platform // URL: <http://www.tomtom.com/lib/doc/licensing/I.LBS.EN.pdf> (дата обращения: 01.05.2017).

55. De Chiara D. From GeoVisualization to visual-analytics: methodologies and techniques for human-information discourse. – 2012.
56. Boy J. D., Uitermark J. How to Study the City on Instagram //PloS one. – 2016. – Т. 11. – №. 6. – С. e0158161.
57. Instagram // URL: <https://www.instagram.com> (дата обращения: 01.05.2017).
58. Eisen M. B. et al. Cluster analysis and display of genome-wide expression patterns //Proceedings of the National Academy of Sciences. – 1998. – Т. 95. – №. 25. – С. 14863-14868.
59. Набор открытых данных «Бесплатные зоны Wi-Fi в общественных местах в Санкт-Петербурге» // URL: [http://data.gov.spb.ru/opendata/7825457753-free\\_wi-fi/](http://data.gov.spb.ru/opendata/7825457753-free_wi-fi/) (дата обращения: 01.05.2017).
60. VK Api method Returns a list of photos. // URL: <https://vk.com/dev/photos.search> (дата обращения: 01.05.2017).
61. Python json module // URL: <https://docs.python.org/2/library/json.html> (дата обращения: 01.05.2017).
62. Hartigan J. A., Wong M. A. Algorithm AS 136: A k-means clustering algorithm //Journal of the Royal Statistical Society. Series C (Applied Statistics). – 1979. – Т. 28. – №. 1. – С. 100-108.
63. Python cluster module // URL: <https://github.com/exhuma/python-cluster> (дата обращения: 01.05.2017).
64. Python PyMongo module // URL: <http://api.mongodb.com/python/current/> (дата обращения: 01.05.2017).
65. PyUnit framework // URL: <https://docs.python.org/2/library/unittest.html> (дата обращения: 01.05.2017).
66. Import plugin repository. URL: <https://bitbucket.org/EduardBlees/importplugin> (дата обращения: 17.04.2017)
67. Intel® Core™ i5-3210M Processor // URL: <https://ark.intel.com/products/67355/> (дата обращения: 01.05.2017).

68. Ubuntu 16.04.2 LTS (Xenial Xerus) // URL: <http://releases.ubuntu.com/16.04/> (дата обращения: 01.05.2017).
69. Damle N., Gray G., Mount B. Curl programming bible. – Wiley, 2002.
70. Cooper M. Advanced Bash Scripting Guide 5.3 Volume 1. – Lulu. com, 2010. – Т. 1.
71. Python time module // URL: <https://docs.python.org/2/library/time.html> (дата обращения: 01.05.2017).
72. Inaba, M.; Katoh, N.; Imai, H. (1994). Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. Proceedings of 10th ACM Symposium on Computational Geometry. pp. 332–339. doi:10.1145/177424.178042

## Приложение А

### Схема архитектуры плагина

На рис. 31 изображена схема архитектуры плагина.

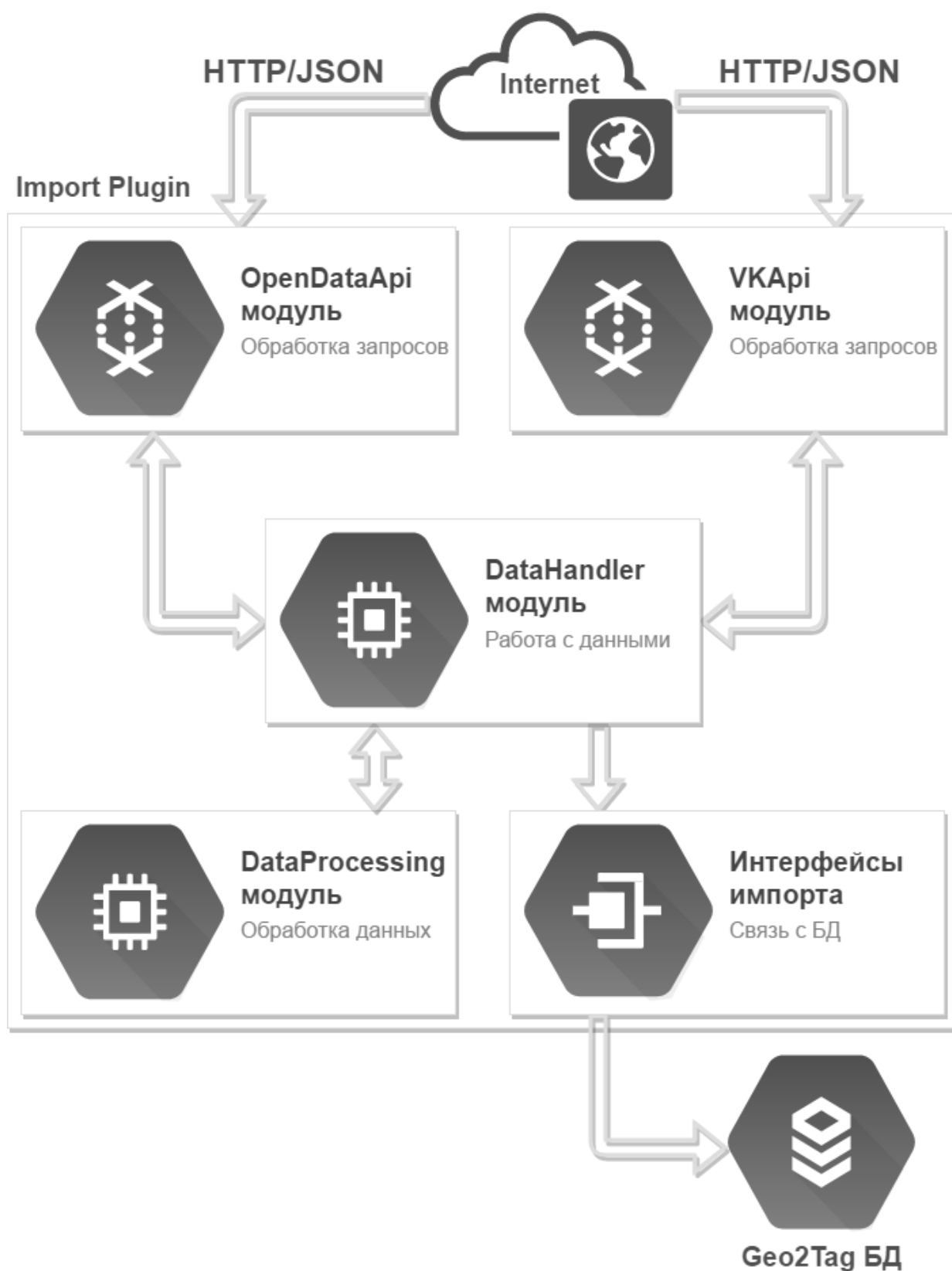


Рисунок 31 – Схема архитектуры плагина

## Приложение Б

### Экспериментально полученные данные

Легенда: Days – промежуток времени в днях, Points – количество анализируемых точек Wi-Fi, FullTime – время работы плагина, OD – время загрузки данных с сайта открытых данных Санкт-Петербурга, VK – время загрузки данных из социальной сети «ВКонтакте», Processing – время обработки данных, Other – «остаток» времени работы плагина – сохранение данных о времени работы в файл и прочее.

Экспериментальные данные исследования зависимости времени работы плагина от промежутка времени выборки (89 точек, 1 – 20 дней) при различных методах обработки данных приведены в табл. 2,3,4.

Таблица 2 – Экспериментальные данные исследования зависимости времени работы плагина от промежутка времени выборки при методе обработки «среднее арифметическое»

Days	FullTime	OD	VK	Processing	Other
1	9,242490905	0,468311071	8,738585949	0,001524925	0,034068959
2	9,639374732	0,479330063	9,141443968	0,001559973	0,017040729
3	9,379333841	0,480347872	8,870051146	0,00139904	0,027535783
4	9,345266837	0,464882135	8,843916893	0,001429081	0,035038728
5	9,71217672	0,489606142	9,194746017	0,001278877	0,026545683
6	9,463733881	0,467215061	8,964991093	0,001221895	0,030305832
7	9,131600806	0,540925026	8,562760115	0,001189947	0,026725718
8	10,42085514	0,66374898	9,730242014	0,001351833	0,025512308
9	10,20510817	0,612741947	9,555365086	0,001391888	0,035609246
10	9,860447392	0,494787931	9,328982115	0,001276016	0,03540133
11	9,610600466	0,496654034	9,080070019	0,001286983	0,03258943
12	10,14635392	0,477252007	9,639408112	0,001376867	0,028316934
13	9,672089544	0,497879028	9,149147034	0,001302004	0,023761478
14	8,873361252	0,475255013	8,369692087	0,001192093	0,027222059
15	9,440755039	0,507876158	8,899283886	0,001341104	0,032253891
16	8,939425637	0,454944134	8,457756996	0,001317024	0,025407483
17	9,818752218	0,458319902	9,327614069	0,000831842	0,031986404
18	9,359334122	0,476535082	8,85819006	0,001281023	0,023327957
19	9,650456581	0,510154963	9,11009407	0,001330853	0,028876696
20	10,26584381	0,468986034	9,77402401	0,000821114	0,022012652

Таблица 3 – Экспериментальные данные исследования зависимости времени работы плагина от промежутка времени выборки при методе обработки «медиана»

Days	FullTime	OD	VK	Processing	Other
1	9,112445832	0,464326143	8,61746192	0,001027107	0,029630662
2	8,996545522	0,473624945	8,496403933	0,001336098	0,025180547
3	10,60416158	0,478984118	10,08773684	0,00141716	0,036023458
4	10,42129322	0,45816803	9,92662096	0,001322031	0,035182201

5	10,79926601	0,469059944	10,29939914	0,001530886	0,02927604
6	9,422062622	0,48987484	8,899112225	0,001430988	0,031644569
7	9,345019577	0,52039814	8,79001689	0,001451969	0,033152578
8	9,199516571	0,510526896	8,656138182	0,000835896	0,032015598
9	9,027141654	0,485184908	8,514444113	0,000823021	0,026689612
10	9,533071899	0,490706921	9,011348963	0,001523972	0,029492044
11	9,512484289	0,451334953	9,02858901	0,001976967	0,030583359
12	9,291049802	0,466960907	8,788324118	0,001271963	0,034492814
13	9,733782921	0,532452822	9,173240185	0,001315832	0,026774082
14	8,031517141	0,478214979	7,517575979	0,001260996	0,034465187
15	9,169816741	0,474019051	8,662213802	0,001280069	0,032303819
16	9,607926298	0,490990162	9,080034971	0,001471996	0,035429169
17	10,10705321	0,472920895	9,603763103	0,001407862	0,028961349
18	9,61770715	0,473795176	9,110400915	0,001568079	0,03194298
19	9,685978935	0,494658947	9,164702892	0,001353025	0,02526407
20	9,491561644	0,533816099	8,93143487	0,001199007	0,025111668

Таблица 4 – Экспериментальные данные исследования зависимости времени работы плагина от промежутка времени выборки при методе обработки «кластеризация»

Days	FullTime	OD	VK	Processing	Other
1	9,551003696	0,565351009	8,939100981	0,020990841	0,025560865
2	9,380730064	0,57923317	8,752584934	0,023428909	0,025483051
3	9,50454485	0,479830027	8,968286037	0,027914993	0,028513793
4	8,698170742	0,454833031	8,193455935	0,019826881	0,030054895
5	8,639628887	0,45216918	8,141242027	0,02461981	0,02159787
6	9,843142078	0,455029964	9,343772888	0,02092909	0,023410135
7	9,649598962	0,618571997	8,979342937	0,026429884	0,025254144
8	9,24695875	0,517323017	8,680476904	0,024240963	0,024917866
9	9,330560644	0,465824842	8,812021017	0,028270952	0,024443832
10	9,841407581	0,486222029	9,303268909	0,019934885	0,031981758
11	10,11401129	0,465653896	9,595664024	0,022651903	0,030041466
12	8,952026651	0,508976936	8,388711214	0,023528807	0,030809694
13	9,7049964	0,470149994	9,179829121	0,022288792	0,032728494
14	9,266203002	0,558589935	8,661346912	0,021661989	0,024604165
15	9,9908224	0,474842072	9,471384048	0,018178942	0,026417338
16	10,86323793	0,529239893	10,2753799	0,026404135	0,032214009
17	10,05394372	0,474173784	9,532106161	0,021908991	0,025754788
18	10,89383645	0,474833012	10,36667204	0,022384159	0,029947235
19	9,401378842	0,527226925	8,813772202	0,027441017	0,032938698
20	9,0127648	0,453669071	8,506995916	0,022726051	0,029373761

Экспериментальные данные исследования зависимости времени работы плагина от количества анализируемых точек (20 дней, 1 – 89 точек) при различных методах обработки данных приведены в табл. 5,6,7.

Таблица 5 – Экспериментальные данные исследования зависимости времени работы плагина от количества анализируемых точек при методе обработки «среднее арифметическое»

Points	FullTime	OD	VK	Processing	Other
1	0,576752522	0,450062037	0,101791143	3,89E-05	0,02486048
2	0,689153989	0,484713078	0,168972969	9,39E-05	0,035374005
3	0,756827247	0,459403992	0,270792961	0,000109911	0,026520383
4	0,811634995	0,474615097	0,309937	8,80E-05	0,026994921
5	1,075916578	0,454062939	0,593829155	0,000100851	0,027923633
6	1,096133973	0,47988987	0,58238101	0,000157118	0,033705975
7	1,216964873	0,44661212	0,74070096	0,000136852	0,029514941
8	1,438516405	0,465322971	0,947965145	0,000149012	0,025079277
9	1,38043757	0,4729321	0,879134893	0,000175953	0,028194623
10	1,606385961	0,456870079	1,126311064	0,000244856	0,022959962
11	1,648715911	0,476100922	1,140120029	0,000197887	0,032297073
12	1,55253913	0,453067064	1,066563845	0,000211954	0,032696267
13	1,980064323	0,505953789	1,447786093	0,000211954	0,026112487
14	2,142710991	0,453152895	1,656918049	0,000236034	0,032404013
15	2,315585064	0,515251875	1,780190945	0,000282049	0,019860195
16	2,049907229	0,506539106	1,512189865	0,000281096	0,030897162
17	2,368177376	0,50910902	1,823889971	0,000257015	0,03492137
18	2,376828626	0,458965778	1,890190125	0,000281096	0,027391628
19	2,545039792	0,497750044	2,019635916	0,000307083	0,027346749
20	2,837193324	0,466057062	2,340296984	0,000309944	0,030529334
21	2,576738499	0,44720006	2,098143101	0,000339031	0,031056307
22	2,59477945	0,449867964	2,112617016	0,000337839	0,031956631
23	2,830869771	0,523798943	2,279213905	0,000347137	0,027509786
24	3,350387099	0,503370047	2,818527937	0,000351906	0,02813721
25	3,46715517	0,492851019	2,941870928	0,000409126	0,032024097
26	3,1894761	0,474031925	2,681208134	0,000501871	0,03373417
27	3,511618122	0,480184078	3,006577969	0,000639915	0,02421616
28	3,298352007	0,475993156	2,787889957	0,000455856	0,034013037
29	3,824414895	0,503389835	3,286020041	0,000449181	0,034555839
30	3,458448626	0,485114098	2,940274954	0,000499964	0,032559611
31	3,587240254	0,471319914	3,086856127	0,00058794	0,028476273
32	4,10005408	0,466834068	3,608191967	0,000473976	0,024554069
33	3,891266883	0,460819006	3,400375128	0,000483036	0,029589713
34	4,0179369	0,456207991	3,525307894	0,00060606	0,035814956
35	4,219541373	0,456827879	3,732347965	0,000544071	0,029821458
36	4,38739419	0,463576078	3,893721819	0,000555038	0,029541254
37	4,864874508	0,502312899	4,332535028	0,000601053	0,029425528
38	4,51720148	0,503415108	3,976279974	0,000597954	0,036908444
39	4,241415034	0,476567984	3,734858036	0,000598907	0,029390107
40	4,50272208	0,491656065	3,981755972	0,00061512	0,028694923
41	5,898702017	0,503533125	5,362462997	0,000659943	0,032045952
42	4,481518641	0,449223995	3,994473934	0,000679016	0,037141695
43	4,804816071	0,472831011	4,298839092	0,000695944	0,032450024
44	5,07896491	0,465047121	4,586174965	0,000669956	0,027072868



## Окончание таблицы 5

45	5,362040147	0,547715902	4,783111095	0,00068593	0,030527219
46	5,132984748	0,480885029	4,629670143	0,000718832	0,021710744
47	5,336921938	0,473374844	4,831766129	0,000716925	0,031064041
48	5,147807682	0,517585039	4,595969915	0,000850916	0,033401812
49	5,765935955	0,481736898	5,253997087	0,000755072	0,029446897
50	5,385930626	0,448523045	4,910691977	0,000762939	0,025952665
51	5,900471022	0,473365784	5,398971081	0,000693083	0,027441075
52	5,891639296	0,473243237	5,389435768	0,000770092	0,028190199
53	6,06348774	0,479697943	5,555614948	0,000757933	0,027416916
54	6,44791183	0,479990005	5,939990997	0,000771046	0,027159782
55	5,957492305	0,474563122	5,45550108	0,00078392	0,026644183
56	6,154708097	0,471031904	5,656011105	0,000803947	0,026861141
57	6,510202677	0,469209909	6,012541056	0,000862837	0,027588875
58	6,659870927	0,504400969	6,12642312	0,000882864	0,028163974
59	6,412445457	0,452163935	5,932142019	0,000833035	0,027306469
60	6,842990094	0,519840956	6,296499014	0,000955105	0,02569502
61	7,023048296	0,467548132	6,523281097	0,000957966	0,031261101
62	6,839999368	0,508852005	6,302078962	0,000895977	0,028172424
63	6,924414283	0,493547916	6,397273064	0,001326799	0,032266504
64	6,481991018	0,479788065	5,968963861	0,001407146	0,031831945
65	7,133397346	0,466122866	6,642444134	0,00105691	0,023773437
66	6,786328753	0,487705946	6,270369053	0,000981092	0,027272662
67	7,251288123	0,479406118	6,739255905	0,000633955	0,031992144
68	7,443067066	0,481550932	6,926695108	0,001013994	0,033807031
69	7,732588852	0,470068932	7,226625919	0,001224995	0,034669006
70	8,55301647	0,738456011	7,782786131	0,001014948	0,03075938
71	8,097650572	0,45699501	7,607926846	0,000967026	0,03176169
72	8,166114651	0,579530954	7,557864189	0,001125813	0,027593695
73	7,868477613	0,608958006	7,225294113	0,001138926	0,033086568
74	10,16385826	0,460980892	9,672626972	0,001585007	0,028665384
75	7,792951271	0,50094986	7,258289099	0,001059055	0,032653257
76	8,021741232	0,491664171	7,499991894	0,001401901	0,028683266
77	8,307075319	0,490061045	7,790110111	0,001580954	0,025323209
78	8,288197988	0,454308987	7,803719044	0,001245975	0,028923982
79	8,679150023	0,479641914	8,170531034	0,001230955	0,027746119
80	8,300863207	0,474177122	7,795567036	0,00125885	0,029860199
81	8,861431047	0,483528137	8,343830824	0,001279116	0,03279297
82	9,136024894	0,55638504	8,552657127	0,001241922	0,025740804
83	9,982528833	0,504719019	9,314763069	0,001163006	0,161883739
84	9,764562844	0,470055819	9,26168704	0,001303911	0,031516074
85	9,645687194	0,450423002	9,170773029	0,00116396	0,023327203
86	9,651901114	0,462681055	9,156594992	0,00147295	0,031152117
87	8,765861902	0,477541924	8,260470152	0,002542019	0,025307808
88	9,569168076	0,462934971	9,079834938	0,001279116	0,025119051
89	9,624985192	0,495646954	9,103291035	0,001104116	0,024943087

Таблица 6 – Экспериментальные данные исследования зависимости времени работы плагина от количества анализируемых точек при методе обработки «медиана»

Points	FullTime	OD	VK	Processing	Other
1	0,673104057	0,517757177	0,122895956	4,51E-05	0,032405862
2	0,729265682	0,492874146	0,210416079	6,29E-05	0,025912515
3	0,792256739	0,47172904	0,295126915	0,00010705	0,025293734
4	1,089289097	0,482225895	0,574754953	9,70E-05	0,032211212
5	1,074796556	0,562758923	0,475977182	0,000113964	0,035946487
6	1,076887374	0,461037159	0,584358931	0,00012207	0,031369214
7	1,3508887	0,454357862	0,869004965	0,000144958	0,027380914
8	1,420691576	0,474498987	0,91219902	0,000234842	0,033758726
9	1,347033438	0,46540904	0,846096039	0,000164986	0,035363373
10	1,521359087	0,456017971	1,029139042	0,000216961	0,035985113
11	1,580084552	0,467453957	1,084156036	0,000200987	0,028273572
12	1,680590311	0,487330914	1,160083055	0,000298023	0,032878319
13	1,810058135	0,486649036	1,295367956	0,000148058	0,027893084
14	1,886742347	0,473651171	1,383136034	0,000154018	0,029801124
15	2,000511695	0,468116045	1,496876955	0,000258923	0,035259772
16	2,102676573	0,470981121	1,606488943	0,00029397	0,024912539
17	2,191893889	0,491499186	1,672310829	0,000257969	0,027825905
18	2,287355827	0,506035089	1,753591776	0,000289202	0,02743976
19	2,663677734	0,45401907	2,176138878	0,000347137	0,033172649
20	2,541604827	0,460613966	2,044571877	0,000498056	0,035920928
21	2,755133626	0,529479027	2,191000938	0,000331163	0,034322497
22	3,544606309	0,492094994	3,022197962	0,000355959	0,029957395
23	2,994250957	0,465986013	2,500931025	0,00037384	0,026960079
24	2,918632162	0,490561008	2,397613049	0,000427008	0,030031097
25	3,094280084	0,47332406	2,588736773	0,000344038	0,031875213
26	3,270521489	0,495580196	2,745626926	0,000417948	0,028896418
27	3,096382796	0,46043396	2,572459936	0,000431061	0,063057839
28	3,083892463	0,470157146	2,583002806	0,000422001	0,03031051
29	3,245117967	0,459085941	2,756558895	0,000465155	0,029007976
30	3,466510503	0,471870899	2,961658001	0,000459194	0,032522409
31	3,619331702	0,492213011	3,098646879	0,000386	0,028085812
32	4,138127101	0,478914022	3,632364988	0,000478029	0,026370061
33	3,944048197	0,47150588	3,435983181	0,000939846	0,03561929
34	3,99625372	0,499712944	3,464270115	0,000531912	0,031738749
35	4,041974529	0,469248056	3,546335936	0,000550032	0,025840505
36	4,048779958	0,468871832	3,554347992	0,000535965	0,025024169
37	4,138815087	0,483092785	3,62905407	0,000607014	0,026061219
38	4,3802327	0,475651026	3,871783972	0,000598907	0,032198795
39	4,378060986	0,470801115	3,872730017	0,00071907	0,033810784
40	4,692980134	0,465383053	4,190377951	0,000624895	0,036594235
41	4,525024778	0,492830038	4,002874851	0,000660181	0,028659708
42	4,6087295	0,673730135	3,904737949	0,000622034	0,029639382
43	4,32560198	0,450731039	3,843288183	0,000631809	0,030950949

## Окончание таблицы 6

44	4,88769976	0,518517971	4,334171057	0,000692129	0,034318603
45	5,045733783	0,462121964	4,554409027	0,00068903	0,028513763
46	5,082850816	0,470832109	4,584218979	0,000672817	0,02712691
47	5,194064124	0,466568947	4,69533205	0,000725985	0,031437142
48	4,827729456	0,488035917	4,306387186	0,000840902	0,03246545
49	5,238488732	0,483838797	4,7286551	0,000818968	0,025175868
50	5,132612203	0,47176981	4,628060102	0,001152992	0,031629299
51	5,20866883	0,457707167	4,721899033	0,000811815	0,028250815
52	5,404308128	0,478534937	4,896982193	0,000816822	0,027974176
53	5,479790476	0,45095396	4,998867035	0,000849962	0,029119518
54	5,532544359	0,461458921	5,034349203	0,000804901	0,035931333
55	5,817526493	0,464533091	5,324354887	0,000883102	0,027755413
56	5,870301989	0,469714165	5,365285873	0,000934124	0,034367827
57	6,042348758	0,483911037	5,529079914	0,000844002	0,028513805
58	6,078417023	0,459042788	5,593641043	0,000833035	0,024900158
59	6,385943744	0,491688967	5,865302086	0,000868797	0,028083894
60	6,245072068	0,492660999	5,72299695	0,000957966	0,028456153
61	6,202696048	0,476956844	5,695364952	0,000936031	0,02943822
62	6,313587998	0,487195969	5,793259144	0,001321077	0,031811808
63	7,328889092	0,456830025	6,835036993	0,000983	0,036039074
64	6,718688881	0,452325821	6,237846136	0,000923872	0,027593052
65	7,09370215	0,466952085	6,592813969	0,000993967	0,032942129
66	6,998624885	0,515784979	6,45730114	0,001015902	0,024522865
67	7,970076491	0,482309818	7,458270073	0,001518011	0,027978589
68	7,231520477	0,476475954	6,729292154	0,00099206	0,024760309
69	7,82843305	0,480145216	7,314245939	0,001102924	0,03293897
70	7,346132106	0,492249966	6,825572968	0,001026154	0,027283019
71	7,695820251	0,464869022	7,199579	0,001080036	0,030292192
72	7,432681483	0,511394024	6,894125938	0,001049995	0,026111525
73	7,597219639	0,486477852	7,083509207	0,001042843	0,026189738
74	7,281287096	0,467353821	6,787479162	0,001292944	0,025161169
75	7,548926418	0,46390295	7,050811052	0,001132011	0,033080404
76	8,137110073	0,455979109	7,647006035	0,001162052	0,032962877
77	7,936388778	0,505226851	7,399369001	0,001113176	0,03067975
78	7,877711657	0,472510099	7,370498896	0,00125289	0,033449772
79	7,843462659	0,469781876	7,343383074	0,001194954	0,029102756
80	8,198388137	0,46022296	7,713855982	0,001213074	0,023096122
81	8,046721989	0,531096935	7,492023945	0,001241922	0,022359186
82	8,605791047	0,47199893	8,106862068	0,001344919	0,02558513
83	9,548794606	0,492539883	9,029651165	0,001198769	0,02540479
84	8,472563394	0,471904993	7,975008011	0,001202106	0,024448284
85	8,920955567	0,473406076	8,418290853	0,001216173	0,028042464
86	9,112550373	0,489394188	8,588657856	0,00130415	0,033194179
87	9,031829689	0,502140045	8,495363951	0,001342058	0,032983635
88	8,993148884	0,469147921	8,496200085	0,001190901	0,026609978
89	9,276933998	0,690401077	8,278183937	0,001307011	0,307041973

Таблица 7 – Экспериментальные данные исследования зависимости времени работы плагина от количества анализируемых точек при методе обработки «кластеризация»

Points	FullTime	OD	VK	Processing	Other
3	0,832363289	0,474980116	0,330224037	0,000339985	0,026819151
4	0,862462201	0,455641031	0,374243975	0,000439167	0,032138028
5	0,97740241	0,468179941	0,474871874	0,00079298	0,033557615
6	1,010844723	0,455737114	0,526085854	0,000652075	0,028369681
7	1,155981987	0,450794935	0,67732501	0,000665903	0,027196138
8	1,3124312	0,457544088	0,823524952	0,000819921	0,030542238
9	1,458011406	0,456237078	0,964864016	0,001004934	0,035905378
10	1,422514716	0,466577053	0,933488846	0,00101614	0,021432677
11	1,60490638	0,537611961	1,032664061	0,001239777	0,033390581
12	1,737254594	0,476497889	1,223762989	0,001332998	0,035660718
13	1,935371289	0,655274153	1,252632856	0,00151515	0,02594913
14	1,940634113	0,479655027	1,426938057	0,00138998	0,032651048
15	2,172965837	0,471690893	1,675408125	0,001795053	0,024071765
16	2,023947368	0,486330032	1,500049114	0,001719952	0,03584827
17	2,131605459	0,485172987	1,613676071	0,001911879	0,030844522
18	2,225543832	0,497129202	1,693488836	0,001863956	0,033061837
19	2,43169837	0,452950001	1,94720006	0,002249002	0,029299307
20	2,326116282	0,471225977	1,825144053	0,002291918	0,027454335
21	2,517187642	0,453052044	2,025832891	0,002490044	0,035812663
22	2,571913595	0,483810902	2,055053949	0,002269983	0,030778761
23	2,737705216	0,480157852	2,21903801	0,002728939	0,035780415
24	2,78312661	0,535479069	2,218185902	0,002793074	0,026668566
25	2,890245187	0,456375122	2,407475948	0,002871037	0,02352308
26	2,874164187	0,463033915	2,373268127	0,003330946	0,034531199
27	3,073472639	0,550112009	2,48536396	0,00432086	0,03367581
28	3,525670322	0,459234953	3,033264875	0,003598213	0,02957228
29	3,751704148	0,454885006	3,267462969	0,003496885	0,025859288
30	3,338409982	0,471590042	2,832527876	0,00504303	0,029249034
31	3,532889752	0,505693913	2,997727156	0,003798008	0,025670676
32	3,658004227	0,515876055	3,104942083	0,003422976	0,033763113
33	4,011954279	0,482419968	3,496134996	0,004554987	0,028844328
34	4,101445988	0,500559092	3,562854052	0,004621983	0,033410862
35	4,039036275	0,470982075	3,531781912	0,005794048	0,03047824
36	4,14071489	0,483274937	3,621287107	0,004731894	0,031420952
37	4,206345039	0,524569035	3,644478083	0,005090952	0,03220697
38	4,173766031	0,475081921	3,664103031	0,004530907	0,030050173
39	4,49627485	0,509457111	3,951894999	0,005757809	0,029164931
40	4,539014946	0,455075979	4,044509888	0,007601976	0,031827103
41	4,472985875	0,491164923	3,951328993	0,005104065	0,025387894
42	4,644254167	0,490578175	4,121490955	0,005748034	0,026437004
43	4,596371856	0,475693941	4,087682009	0,005888939	0,027106967
44	4,902972274	0,469535112	4,401268959	0,006607056	0,025561147
45	4,487876702	0,465284109	3,983816862	0,009319067	0,029456664
46	4,77376789	0,497615814	4,244382143	0,006263971	0,025505961

47	5,151259468	0,471561909	4,649010181	0,007401943	0,023285435
48	5,024352525	0,524316788	4,461429119	0,006649017	0,031957601
49	5,29396138	0,483332872	4,777408123	0,007496834	0,025723551
50	4,969179866	0,465052843	4,460460186	0,007777929	0,035888908
51	5,423415799	0,474573851	4,912460089	0,007805824	0,028576035
52	5,640935459	0,497354031	5,108483076	0,008262873	0,02683548
53	5,527031252	0,471163988	5,014431953	0,008061171	0,03337414
54	5,624588852	0,501431942	5,08554101	0,007721901	0,029893999
55	5,594680728	0,468885899	5,088096142	0,008111954	0,029586734
56	5,679413838	0,495585918	5,142369032	0,012159109	0,029299779
57	5,782173377	0,510859013	5,240127087	0,009483099	0,021704179
58	6,157451872	0,45984602	5,658812046	0,009120941	0,029672865
59	5,936136266	0,459844828	5,432964087	0,013336897	0,029990455
60	6,388910563	0,457967997	5,885233164	0,018701792	0,027007611
61	6,286156709	0,473494053	5,766678095	0,009929895	0,036054666
62	6,318356155	0,474875212	5,799244881	0,010509968	0,033726095
63	6,243268259	0,476030111	5,724103928	0,012650967	0,030483253
64	6,47455232	0,48520112	5,948427916	0,011061907	0,029861377
65	6,836020951	0,46417284	6,325878143	0,011194944	0,034775023
66	6,951678805	0,45351696	6,451912165	0,010914803	0,035334878
67	6,902353637	0,526542902	6,336291075	0,012356043	0,027163617
68	6,638867724	0,464517117	6,134245872	0,01285696	0,027247775
69	6,777607073	0,464816093	6,269490957	0,009557009	0,033743014
70	6,914033322	0,465413094	6,40546298	0,011904001	0,031253247
71	7,087599934	0,464421034	6,579487085	0,013608932	0,030082882
72	7,305299552	0,455569029	6,804069996	0,013312101	0,032348426
73	7,45110597	0,448568821	6,959812164	0,013542891	0,029182094
74	7,273865682	0,475141048	6,75667882	0,01312995	0,028915864
75	7,790346084	0,472994089	7,272387981	0,014554977	0,030409036
76	7,514476544	0,509855986	6,957561016	0,014010906	0,033048636
77	8,019355958	0,453493834	7,520430088	0,019696951	0,025735086
78	7,724902075	0,458055973	7,234184027	0,017323971	0,015338104
79	8,484789559	0,472769976	7,974575996	0,009057999	0,028385588
80	8,150556492	0,475537062	7,632531881	0,014842033	0,027645516
81	8,320458626	0,448472977	7,828441858	0,016561031	0,02698276
82	8,218140368	0,475204945	7,696830034	0,015734911	0,030370478
83	9,404927089	0,477452993	8,878674984	0,02196908	0,026830032
84	8,613301368	0,473031044	8,089055061	0,016808033	0,03440723
85	9,000845653	0,520951986	8,438202143	0,016752958	0,024938566
86	9,59778954	0,482843161	9,06880784	0,020697117	0,025441422
87	8,903419129	0,453659058	8,401633978	0,019307852	0,028818242
88	9,352770719	0,466280937	8,832985878	0,026354036	0,027149867
89	9,280558339	0,477034807	8,759245157	0,018565855	0,02571252