

Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
Направление подготовки 09.04.02 «Информационные системы и технологии»  
Отделение информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Алгоритмическое и программное обеспечение поиска и анализа целевых сообществ в социальной сети

УДК 004.415.2:004.421:004.738

**Студент**

Группа	ФИО	Подпись	Дата
8ИМ82	Кузнецов Сергей Анатольевич		

**Руководитель**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Конотопский Владимир Юрьевич	к. э. н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Горбенко Михаил Владимирович	к. т. н.		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
09.04.02 "Информационные системы и технологии", профиль "Геоинформационные системы"	Шерстнев Владислав Станиславович	к. т. н., доцент		

Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа: Инженерная школа информационных технологий и робототехники  
Направление подготовки (специальность): 09.04.02 Информационные системы и технологии  
Отделение школы (НОЦ): Отделение информационных технологий

УТВЕРЖДАЮ:  
Руководитель ООП

\_\_\_\_\_  
(Подпись)      (Дата)      (Ф.И.О.)

### ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации
--------------------------

Студенту:

Группа	ФИО
8ИМ82	Кузнецов Сергей Анатольевич

Тема работы:

Алгоритмическое и программное обеспечение поиска и анализа целевых сообществ в социальной сети	
Утверждена приказом директора (дата, номер)	28.02.2020г., №59-47/с

Срок сдачи студентом выполненной работы:	04.06.2020г.
--	--------------

### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p><b>Исходные данные к работе</b></p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<ol style="list-style-type: none"> <li>1. Авторизованный доступ к программному интерфейсу приложений социальной сети «ВКонтакте».</li> <li>2. База знаний, описывающая ключевые слова и даты целевых сообществ.</li> </ol>
---	--

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> <li>1. Анализ научной и технической документации, описывающей предметную область, а также существующие подходы и решения в области поиска целевых сообществ социальной сети.</li> <li>2. Разработка архитектуры программного обеспечения поиска и анализа целевых сообществ социальной сети.</li> <li>3. Разработка алгоритмов параметризованного поиска целевых сообществ.</li> <li>4. Разработка алгоритмов визуализации геолокации участников сообществ социальной сети.</li> <li>5. Программная реализация алгоритмов.</li> <li>6. Проектирование и реализация пользовательского интерфейса.</li> <li>7. Оценка конкурентоспособности разработки, расчет затрат на проведение исследования.</li> <li>8. Анализ условий труда исполнителей проекта.</li> </ol>
<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> <li>1. еЕРС-диаграммы разрабатываемых алгоритмов.</li> <li>2. Схема высокоуровневой архитектуры разрабатываемого программного обеспечения.</li> </ol>
<p><b>Консультанты по разделам выпускной квалификационной работы</b></p> <p><i>(с указанием разделов)</i></p>	
<p><b>Раздел</b></p>	<p><b>Консультант</b></p>
<p>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Доцент, к.э.н., Конотопский Владимир Юрьевич</p>
<p>Социальная ответственность</p>	<p>Доцент, к.т.н., Горбенко Михаил Владимирович</p>
<p>Раздел на иностранном языке</p>	<p>Старший преподаватель ОИЯ ШБИП, Маркова Наталия Александровна</p>
<p><b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b></p>	
<p>Введение</p>	
<p>Аналитический обзор научной и технической документации</p>	
<p><b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b></p>	<p>15.01.2020г.</p>

**Задание выдал руководитель / консультант (при наличии):**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	К.Т.Н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ИМ82	Кузнецов Сергей Анатольевич		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

**Инженерная школа информационных технологий и робототехники**  
 Направление подготовки 09.04.02 – Информационные системы и технологии  
 Уровень образования: магистр  
 Отделение информационных технологий  
 Период выполнения: весенний семестр 2020 учебного года

Форма представления работы:

Магистерская диссертация
--------------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	04.06.2020г.
--	--------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
10.04.2020	Раздел 1. Аналитический обзор научной и технической документации	20
15.04.2020	Раздел 2. Разработка архитектуры и алгоритмического обеспечения для поиска и анализа целевых сообществ	30
30.04.2020	Раздел 3. Реализация программного обеспечения для поиска и анализа целевых сообществ	30
30.05.2020	Раздел 4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.	10
03.06.2020	Раздел 5. Социальная ответственность	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Шерстнев Владислав Станиславович	к.т.н.		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ИМ82	Кузнецов Сергей Анатольевич

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	Информационные системы и технологии

### Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Использовать действующие ценники и договорные цены на потребленные материальные и информационные ресурсы, а также указанную в МУ величину тарифа на эл. энергию
2. Нормы и нормативы расходования ресурсов	—
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Действующие ставки единого социального налога и НДС, ставка дисконтирования = 0,1 (см. МУ)

### Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Дать характеристику существующих и потенциальных потребителей (покупателей) результатов ВКР, ожидаемых масштабов их использования
2. Разработка устава научно-технического проекта	Разработать проект такого устава в случае, если для реализации результатов ВКР необходимо создание отдельной организации или отдельного структурного подразделения внутри существующей организации
3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Построение плана-графика выполнения ВКР, составление соответствующей сметы затрат, расчет цены результата ВКР.
4. Определение ресурсной, финансовой, экономической эффективности	Оценка экономической эффективности использования результатов ВКР, характеристика других видов эффекта

### Перечень графического материала (с точным указанием обязательных чертежей):

1. «Портрет» потребителя результатов НТИ
2. Сегментирование рынка
3. Оценка конкурентоспособности технических решений
4. Диаграмма FAST
5. Матрица SWOT
6. График проведения и бюджет НТИ - выполнить
7. Оценка ресурсной, финансовой и экономической эффективности НТИ - выполнить
8. Потенциальные риски

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский В.Ю.	к.э.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ИМ82	Кузнецов С.А.		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ИМ82	Кузнецов Сергей Анатольевич

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	Информационные системы и технологии

Тема ВКР:

<b>Алгоритмическое и программное обеспечение поиска и анализа целевых сообществ в социальной сети</b>	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Программное обеспечение, которое применяется для поиска и анализа сообществ в социальной сети с заданными характеристиками.
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<ul style="list-style-type: none"> <li>– СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы</li> <li>– ГОСТ 12.2.032-78 ССБТ Рабочее место при выполнении работ сидя. Общие эргономические требования</li> <li>– ГОСТ Р 50923-96. Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения</li> </ul>
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<ul style="list-style-type: none"> <li>– Отклонение показателей микроклимата.</li> <li>– Недостаточная освещённость рабочей зоны.</li> <li>– Превышение уровня шума.</li> <li>– Повышенный уровень электромагнитных излучений.</li> <li>– Электробезопасность.</li> <li>– Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на работника.</li> </ul>
<b>3. Экологическая безопасность:</b>	Воздействие на окружающую среду вредными и опасными отходами.
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Предотвращение возникновения пожароопасной ситуации

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД	Горбенко Михаил Владимирович	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ИМ82	Кузнецов Сергей Анатольевич		

### Планируемые результаты обучения

Код	Результаты обучения
<b>09.04.02 «Информационные системы и технологии» (Геоинформационные технологии)</b>	
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и представления информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.
P3	Демонстрировать способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Демонстрировать способность к самостоятельному обучению новым методам исследования, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения.
P5	Владеть современными коммуникативными технологиями, в том числе на иностранном языке для академического и профессионального взаимодействия.
P6	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современных программных и информационных систем, в управлении коллективом. Способность организовывать и эффективно руководить работой команды проекта при разработке программных и информационных систем.
P7	Разрабатывать стратегии проектирования, критерии эффективности и ограничения применимости новых методов и средств проектирования и разработки программных систем.
P8	Планировать и проводить теоретические и экспериментальные (численные) исследования в области создания программных систем. Оценивать и выбирать вариант архитектуры информационной системы.
P9	Владеть методами и средствами обработки и анализа пространственных данных, методами и средствами разработки хранилищ данных и программного обеспечения геоинформационных систем и технологий.
P10	Владеть современными инструментальными CASE-средствами программирования и технологиями управления и обработки данными. Использовать их при разработке требований, проектировании и создании программного обеспечения геоинформационных систем и технологий.
P11	Осуществлять проектирование и разработку геоинформационных систем и технологий в среде корпоративных и глобальных информационно-телекоммуникационных систем.
P12	Осуществлять управление процессами внедрения/сопровождения (модернизации, интеграции) информационных, геоинформационных систем и технологий.



## РЕФЕРАТ

Выпускная квалификационная работа 102 с., 15 рис., 17 табл., 34 источника, 2 прил.

Ключевые слова: социальная сеть, целевое сообщество, поиск, анализ, база знаний.

Объектом исследования являются сообщества пользователей социальной сети ВКонтакте.

Цель работы – разработка алгоритмического и программного обеспечения поиска и анализа целевых сообществ в социальной сети.

В процессе исследования был проведен анализ разработанных алгоритмов поиска и анализа сообществ в социальной сети. Были разработаны новые алгоритмы поиска и анализа целевых сообществ и интерфейс приложения.

В результате исследования были формализованы требования к алгоритмическому и программному обеспечению. Спроектирована высокоуровневая архитектура системы поиска и анализа целевых сообществ в социальной сети, разработаны алгоритмы поиска и анализа сообществ. Разработан интерфейс приложения.

Степень внедрения: данная работа проводилась в рамках грантов РФФИ и ВИУ, по результатам выполненной работы оформлен акт о внедрении.

Область применения: социо-политические исследования, маркетинговые исследования.

Экономическая эффективность/значимость работы: данная разработка имеет высокую окупаемость / обеспечивает автоматизацию процесса поиска и анализа целевых сообществ в социальной сети для целей социо-политических исследований. В настоящий момент специализированное программное обеспечение для таких целей отсутствует.

В будущем планируется разработка дополнительных алгоритмов анализа целевых сообществ и расширение базы знаний.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	14
1 Аналитический обзор научной и технической документации .....	15
1.1 Методы и решения для поиска целевых сообществ в социальных сетях..	15
1.2 Особенности и ограничения процесса анализа сообществ в социальных сетях.....	18
1.3 Спецификация требований к разрабатываемому алгоритмическому и программному обеспечению .....	22
1.4 Выводы по аналитическому обзору научной и технической документации .....	25
2 Разработка архитектуры и алгоритмического обеспечения для поиска и анализа целевых сообществ .....	26
2.1 Высокоуровневая архитектура программного обеспечения .....	26
2.1.1 Получение данных из социальной сети.....	27
2.1.2 Хранилище данных.....	28
2.1.3 База знаний .....	29
2.2 Алгоритмическое обеспечение для поиска целевых сообществ.....	30
2.2.1 Алгоритм поиска по ключевым словам.....	30
2.2.2 Алгоритм поиска смежных сообществ .....	31
2.3 Алгоритмическое обеспечения анализа целевых сообществ .....	33
2.3.1 Алгоритмы формирования «профиля» группы .....	33
2.3.2 ККА или алгоритм анализа календарной активности сообщества.....	36
2.3.3 Визуализация геолокационных данных .....	37
3 Реализация программного обеспечения для поиска и анализа целевых сообществ .....	39
3.1 Краткое описание используемых технологий.....	39
3.1.1 Язык программирования Python 3.....	39
3.1.2 Фреймворк Django.....	39
3.1.3 HTML .....	40
3.1.4 CSS.....	40
3.1.5 BootStrap .....	41

3.1.6 База данных Mongo.....	42
3.1.7 Система контроля версий git и облачный репозиторий github.com.....	42
3.2 Программная реализация пользовательского интерфейса .....	44
3.3 Апробация программного обеспечения.....	47
3.4 Заключение по разделу .....	50
4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	51
4.1 Организация и планирование работ.....	51
4.1.1 Продолжительность этапов работ .....	52
4.2 Расчет сметы затрат на выполнение НИР .....	56
4.2.1 Расчет затрат на материалы .....	57
4.2.2 Расчет заработной платы .....	57
4.2.3 Расчет затрат на социальный налог .....	59
4.2.4 Расчет затрат на электроэнергию .....	59
4.2.5 Расчет амортизационных расходов.....	60
4.2.6 Расчет прочих расходов.....	61
4.2.7 Расчет общей себестоимости разработки .....	61
4.2.8 Расчет прибыли .....	62
4.2.9 Расчет НДС .....	62
4.2.10 Цена разработки проекта.....	62
4.3 Оценка экономической эффективности проекта.....	62
4.3.1 Определение срока окупаемости инвестиций (PP – payback period)....	63
4.4 Заключение по разделу .....	67
5 Социальная ответственность .....	68
5.1 Правовые и организационные вопросы обеспечения безопасности.....	68
5.2 Производственная безопасность.....	72
5.2.1 Микроклимат рабочего места .....	74
5.2.2 Освещенность рабочей зоны .....	75
5.2.3 Шум .....	78
5.2.4 Электромагнитное излучение.....	79
5.2.5 Опасность поражения током.....	80
5.3 Экологическая безопасность.....	83

5.4 Безопасность в чрезвычайных ситуациях .....	85
5.5 Заключение по разделу .....	86
Список используемых источников:.....	87
Приложение А .....	91
Приложение Б .....	104

### Перечень условных обозначений

БК	Вконтакте
SMM	Social Media Marketing
API	Application Programming Interface
ВУЗ	Высшее учебное заведение
KW&KD	KeyWords & KeyDates
ККА	Календарно-корреляционный анализ
ID	Идентификатор пользователя или сообщества в социальной сети
MVC	Model View Controller
DRY	Don't repeat yourself
HTML	HyperText Markup Language
SGML	Standard Generalized Markup Language
XHTML	extensible hypertext markup language
DOM	Document Object Model
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
CSS	Cascading Style Sheets
XML	eXtensible Markup Language
SVG	Scalable Vector Graphics
XUL	XML User Interface Language
ВИУ	Ведущий исследовательский университет
РФФИ	Российский фонд фундаментальных исследований

## **ВВЕДЕНИЕ**

В 2018 году были проведены регулярные исследования активной аудитории социальных сетей в России.

Самая распространённая социальная сеть в России ВКонтакте: сообщений в месяц – 1 096 392 000, авторов в месяц – 36 453 000.

Основная цель социальных сетей – увеличить качество потребления медиа информации и удержать пользователя. Предложить новые возможности заработка авторам и поставщикам маркетингового контента. Конкуренция между социальными сетями за пользователей играет на руку и пользователям, и качеству представляемой информации.

Исходя из этого, можно сделать вывод о том, что социальные сети и сервисы приобретают всё больший вес как источник исходных данных для социологических, политологических и иных исследований аудитории.

Целью работы является разработка алгоритмического и программного обеспечения для повышения эффективности поиска и анализа целевых сообществ в социальных сетях для целей социо-политических исследований.

В первом разделе проведён аналитический обзор, описана актуальность работы и сформированы требования к разрабатываемому алгоритмическому и программному обеспечению.

Второй раздел посвящён проектированию и программной реализации алгоритмического обеспечения.

Третий раздел посвящён проектированию и программной реализации интерфейса приложения и апробации программного обеспечения.

В качестве площадки для реализации и апробации разработанного алгоритмического и программного обеспечения была использована социальная сеть «ВКонтакте».

## **1 Аналитический обзор научной и технической документации**

### **1.1 Методы и решения для поиска целевых сообществ в социальных сетях**

Работы российских авторов в области анализа социальных сетей посвящены описанию основных компонентов разработанного стека технологий для анализа пользовательских данных из социальной сети [1], исследованию популярных практических инструментов и сервисов анализа данных о поведении целевой аудитории в Интернете [2], проблеме автоматизации сбора и аналитики целевой аудитории в социальных сетях на примере наиболее функционального на настоящий момент программного решения для SMM – программы «Церебро Таргет» [3], представлению результатов Томского государственного университета по формированию модели абитуриента на основе данных из соцсети ВКонтакте и разработке инструментов выгрузки и анализа данных и поиску одарённых абитуриентов для поступления [4], предложена Data Mining - концепция выявления сообществ потенциальных абитуриентов в социальных сетях и их лидеров, реализовано программное приложение, выполняющее сбор информации из социальных сетей в рамках построения экспериментальной модели поиска сообществ и их лидеров [5].

Авторы [2, 3] для исследований используют существующие парсинговые сервисы, авторы [1, 4, 5] разработали свои инструменты с использованием программного интерфейса приложения API ВК.

В исследовании [4] была проверена гипотеза о взаимосвязи интересов школьников в соцсетях и выбору направления обучения в вузе. Профиль школьника смоделирован по простой классификации: гуманитарные, точные и естественные науки. Для гуманитариев точность составила 0,82, для точных наук - 0,76, для естественных - 0,69. Также подтверждена гипотеза, что контент страницы школьника может отражать научные и профессиональные интересы.

Исследования [1, 4, 5] для поиска целевых сообществ используют информацию о социальных связях пользователей – список друзей.

Интерес представляют методы и алгоритмы поиска целевых сообществ предложенный в [5]. Авторы предлагают рассматривать совокупность поступающих школьников как взаимосвязанную сложную неоднородную структуру, с лидерами, динамикой и явными и неявными связями.

Такой взгляд актуализирует необходимость разработки концепции информационной системы, позволяющей выполнять поиск, накопление информации из социальной сети об абитуриентах, анализировать связи между ними, находить их сообщества и лидеров. Предложенная система является маркетинговым решением, состоящая из этапа сбора, предобработки, анализа и представления полученных результатов.

Исходная точка для сбора информации и построения модели – показатели географического положения ВУЗа и задание радиуса поиска. Данные, собираемые из соцсети, имеют характерные особенности – высокую размерность, большой объём и наличие пропусков. Поэтому возникает задача их обоснования, отбора и редукции перед началом построения моделей. Проблема решается путём применения последовательной редукции исходных данных – выделение целевых групп пользователей соцсети с последующей кластеризацией графа связей между отобранными пользователями (поиск явных и неявных сообществ). В результате выделяются группы наиболее вероятных абитуриентов, внутри которых происходит поиск лидеров (наиболее влиятельных агентов).

Для выявления наиболее активных участников целевого сообщества с точки зрения их информационного взаимодействия с другими пользователями для определённых целевых групп строят граф взаимосвязей.

Выявлению лидеров сообществ абитуриентов происходит путём расчёта значения Betweenness Centrality для каждой вершины. Данный алгоритм при поиске лидеров сообществ требует задания порогового значения, при котором вершины, имеющие такую и большую величину, идентифицируются как лидеры.

В данной статье [5] обоснована Data Mining-концепция системы обнаружения сообществ потенциальных абитуриентов и их лидеров на основе



информационного подхода, которая, в отличие от известных, позволяет автоматически проводить сбор и анализ информации об абитуриентах из соцсетей, строить модели их скрытых сообществ, выявлять агентов влияния с целью повышения эффективности проведения агитационных мероприятий.

Таким образом, основная часть исследований, представленных в русскоязычном поле, направлена на использование в качестве источника данных уже разработанных парсинговых сервисов и лишь малая доля исследований касается непосредственно процессов сбора, предварительной обработки, анализа и интерпретации данных. При этом, исследования, которые включают полный перечень задач, касаются узкоспециализированных задач, в частности, подбора абитуриентов для поступления в университет.

Работы в англоязычных источниках в области анализа социальных сетей посвящены разработке подхода для измерения социальной напряженности в отдельных регионах России, основанного на анализе сообщений пользователей в социальной сети ВК [6], обсуждению методов и алгоритмов выявления неявных сообществ – виртуальных сообществ со схожими интересами [7], созданию руководства по проблеме обнаружения сообщества [8], методам обнаружения крупных динамичных онлайн-активистов и получения нового представления о крупных интернет-активистах [9], поиску сообществ с использованием методов анализа графа социальных связей [10], разработке метода поиска сообщества с использованием измерения влияния [11], разработке алгоритма для поиска подграфов с заданными свойствами в больших социальных сетях [12], предложению простой и новой структуры для обнаружения сообществ на основе топологии сети [13], разработке простого и эффективного подхода, который одновременно раскрывает сообщества и их центры [14].

Таким образом, основная часть исследований, представленных в англоязычном поле, направлена на использование информации социальных связях пользователей для поиска целевых сообществ.

В результате аналитического обзора можно сделать вывод о том, что проблему повышения точности поиска целевых сообществ нужно рассматривать применительно к конкретным задачам, а не в общем. Существует необходимость разработки методов повышения релевантности поиска целевых сообществ в социальной сети, использующих характерные признаки сообществ.

## **1.2 Особенности и ограничения процесса анализа сообществ в социальных сетях**

Один из наиболее часто используемых методов - получение данных через интерфейсы прикладного программирования (API) из сайтов социальных медиа. Использование API очень удобно, но существует ограничение - социальная сеть отдаёт фиксированный объём информации за определённое время, что может создать сложность при получении информации. Также социальная сеть отдаёт не все данные, которые отображаются в интерфейсе. Это связано с тем, что социальные сети сохраняют приватность своих пользователей, а также снижает нагрузку на серверную часть.

Не зная статистических характеристик пользователей, мы не можем гарантировать, что образцы данных, полученные из социальной сети являются репрезентативными и, следовательно, результаты анализа образцов данных являются достоверными.

В классической литературе интеллектуального анализа данных, успешное получение данных влечет за собой обширную предварительную обработку данных и удаление шума [15]. По своей природе данные социальных сетей могут содержать большую часть шума. Для этих данных можно отметить следующее: слепое удаление шума может усугубить проблему, потому что удаление может также устранить важную информацию, и определение шума становится сложным, потому что зависит от задачи.

Стандартная процедура оценки закономерностей в полученных данных должна иметь практическую пользу. Например, набор данных может быть разделен на учебные и тестовые наборы. Только обучающие данные используются в обучении, а тестовые данные служат основной истиной для

тестирования. Однако практическая польза часто недоступна в полученных данных из социальных сетей.

Объем данных в социальных сетях огромен. Тем не менее, когда мы приближаемся к конкретному пользователю, для которого, например, мы хотели бы сделать соответствующие выводы, мы часто имеем мало данных. Мы должны использовать характеристики социальных сетей и использовать их многомерные, многоисточниковые и многоузловые данные для агрегирования информации с достаточной статистикой для эффективного получения данных.

Социальные сети дают возможность новым коллективам исследователей изучать человеческие отношения [16]. Эти большие сети создают много новых проблем при получении данных.

Некоторые примеры приведены ниже.

Масштабируемость. Сеть пользователей, представленная в социальных сетях, может быть огромной и включать миллионы связей, в то время как традиционная социальная сеть имеет дело с сотнями человек и меньше. Существующие методы анализа традиционной сети могут давать ошибку при применении непосредственно к сетям такого громадного размера.

Гетерогенность. В реальности между людьми может существовать множество отношений. Два человека могут быть одновременно друзьями и коллегами. Между тем же набором пользователей в сети существует множество взаимоотношений. В одну сеть могут быть вовлечены разные типы объектов.

В одной сети находятся разнородные объекты, так как во многих сайтах обмена мультимедийной информации пользователи, теги и контент переплетены. Анализ этих гетерогенных сетей, включающих гетерогенные объекты или взаимодействия, требуют новых теорий и инструментов.

Эволюция. Существует отличие от классического веб-майнинга. На сайтах обмена контентом и в блогосфере люди быстро теряют интерес к общему контенту и сообщениям в блоге. Присоединяются новые пользователи, между существующими членами устанавливаются новые связи, давние пользователи становятся неактивными или просто выходят из социальной сети.

Коллективный разум. Пользователи в социальных сетях часто делятся своими связями. Мнение сообщества, в виде комментариев, рейтингов и обзоров доступны всем. Данный контент в сочетании с взаимосвязями пользователей может быть полезен для многих исследований. Эффективное извлечение информации о социальных связях и мнений сообщества при создании приложений для социальных сетей по-прежнему остается проблемой.

Оценка. При майнинге социальных сетей, оценка является исследовательским барьером. В традиционном интеллектуальном анализе данных мы привыкли к учебно-тестовой модели оценки. Так как многие социальные сети обязаны защищать информацию о конфиденциальности пользователей, оценка отличается в социальных сетях. Также отсутствие достоверной информации для многих исследований социальных сетей ещё более затрудняет некоторое сравнительное изучение - другая часто встречающаяся проблема.

К основным задачам анализа социальных сетей можно отнести:

1) Моделирование сети

Сетевое моделирование предлагает глубокое понимание динамики сети, которая не зависит от сетевых доменов. Модель сети может использоваться для исследования различных свойств сети, например, распространение информации в данной структуре сети, устойчивость атакуемой сети и т. д.

Когда сети увеличиваются до миллионов узлов, проблемой становится вычисление статистики сети, такой как средний коэффициент кластеризации и диаметр.

В последнее время все большее внимание привлекают методы использование возможностей распределенных вычислений.

2) Анализ центральности и моделирование влияния

Анализ центральности заключается в определении наиболее «важных» узлов в сети.

Традиционный анализ социальных сетей зависит от структуры ссылок для идентификации узлов с высокой центральностью. Обычно используется критерий центральности.

Дополнительная информация доступна в социальных сетях, такая как комментарии и теги. Для изучения центральности эта информация открывает возможности для объединения различных источников информации.

Смежной задачей является исследование влияния, направленное на понимание процесса влияния. Множество узлов, которые увеличивают влияние на население.

### 3) Обнаружение сообществ

В разных контекстах сообщества также называются группами, связными подгруппами, модулями или кластерами. Это одна из фундаментальных задач в анализе социальных сетей. «Основатели» социологии утверждали что причины социальных явлений должны были быть найдены изучением сообществ, а не отдельных индивидуумов. Определить набор узлов, как они взаимодействуют друг с другом - найти сообщество в социальной сети.

Обнаружение сообщества может облегчить другие задачи социальных сетей и применяться в других важных приложениях. Например, группировка пользователей со схожими интересами в социальной сети предоставляет эффективные предложения, которые открывают продавцам широкий спектр соответствующих товаров для улучшения показателей продаж. Сообщества также могут использоваться для уменьшения большой сети в гораздо меньшую сеть. Таким образом, решение проблемы осуществлять не на уровне сети, а на уровне группы.

### 4) Классификация и рекомендация

Для социальных сетей распространены задачи классификации и рекомендации. У большого количества пользователей хороший сайт социальных сетей часто востребован. Для сайта стимулировать больше взаимодействий пользователей очень важно. Для этого предоставляются индивидуальные рекомендации (например, сообщества, друзья, теги).

Классификация может помочь рекомендации. Например, рекомендации друзей, в которых предлагается список друзей, которые может знать пользователь - одна общая черта во многих социальных сетях.

Это эффективный способ расширить связи пользователя, так как пользователю трудно узнать, кто зарегистрирован в социальной сети.

#### 5) Конфиденциальность, спам и безопасность

Социальные сети влекут за собой тесный контакт между людьми.

Многие сайты социальных сетей (например, Facebook, Google Buzz) часто находятся в спорах и разбирательствах по вопросу конфиденциальности пользователей. Это та задача, которую не легко решить.

Спам - еще одна проблема, вызывающая серьезную обеспокоенность в социальных сетях.

СМИ часто публикуют большой объем личной информации.

Некоторые спамеры для получения частной информации пользователей в социальных сетях используют поддельные аккаунты.

Таким образом, майнинг в социальных сетях - это молодое и динамичное направление с большими возможностями. Социальные сети удивляют нас своими новыми формами и разнообразием. Текущая тенденция заключается в том, что социальные сети все больше вписывается в общество с распространением мобильных технологий. Другими словами, пропасть между виртуальным и физическим миром сокращается, поскольку социальные сети все больше и больше входят в жизнь людей, что ведёт к сближению алгоритмов анализа данных социальных сетей и алгоритмов анализа реальных данных.

### **1.3 Спецификация требований к разрабатываемому алгоритмическому и программному обеспечению**

Основные требования к разрабатываемому программному обеспечению:

- Должна быть функция автоматизированного поиска целевых сообществ по ключевым словам и смежным сообществам.
- Должна быть функция поиска связей и взаимного влияния сообществ.

- Должна быть функция оценки календарной активности целевого сообщества.
- Должна быть функция визуализации связей сообществ и статистической информации.

Описание основных интерфейсов и вариантов использования

#### *Требования к авторизации*

На странице авторизации пользователю должны быть доступны:

- Ввод значений логина и соответствующего пароля.
- Переход на страницу регистрации нового пользователя.
- Переход на страницу восстановления пароля.

#### *Требования к регистрации*

На странице регистрации пользователь должен заполнить обязательные поля: имя учетной записи, пароль (и подтверждение пароля), контактный email.

#### *Требования к Странице по умолчанию*

Должна содержать общее описание возможностей приложения

#### *Поиск*

##### *Требования к Поиску - KW&KD (ключевые слова и даты)*

На странице авторизованному пользователю должны быть доступны следующие действия:

- Ручной ввод значений ключевых слов и словосочетаний для поиска целевых сообществ.
- Выбор ранее созданного набора KW&KD в качестве исходных данных для поиска сообществ.
- Импорт excel-файла с набором KW&KD в качестве исходных данных.
- Задание значений дополнительных фильтров результатов поиска (число пользователей сообщества, даты активности).
- Редактирование результатов поиска.
- Сохранение результатов поиска в качестве нового списка сообществ.

##### *Требования к Поиску смежных сообществ*

На странице авторизованному пользователю должны быть доступны следующие действия:

- Ручной ввод идентификаторов сообществ для поиска смежных сообществ.
- Выбор ранее созданного списка сообществ в качестве исходных данных для поиска смежных.
- Импорт excel-файла со списком сообществ в качестве исходных данных.
- Задание значений дополнительных фильтров результатов поиска (число пользователей сообщества, даты активности).
- Редактирование результатов поиска.
- Сохранение результатов поиска в качестве нового списка сообществ.

#### *База знаний*

##### *Требования к Базе знаний - KW&KD*

На странице авторизованному пользователю должны быть доступны следующие действия:

- Создание нового набора KW&KD с ручным вводом данных.
- Создание нового набора KW&KD на основе уже имеющегося.
- Создание нового набора KW&KD на основе импортируемого excel-файла.
- Редактирование выбранного набора KW&KD.
- Экспорт выбранного набора KW&KD в excel-файл.

##### *Требования к Базе знаний - Сообщества*

На странице авторизованному пользователю должны быть доступны следующие действия:

- Создание нового списка сообществ с ручным вводом данных.
- Создание нового списка сообществ на основе уже имеющегося.
- Создание нового набора списка сообществ на основе импортируемого excel-файла.
- Редактирование выбранного списка сообществ.
- Экспорт выбранного набора списка сообществ в excel-файл.

#### *Анализ*



### *Требования к Анализу – ККА (Календарно-корреляционной анализ)*

На странице авторизованному пользователю должны быть доступны:

а) ввод исходных данных для анализа:

- списка сообществ:
  - сформированного вручную;
  - импортированного из excel-файла;
  - ранее созданного.
- ранее созданного набора KW&KD.

б) создание задачи для выполнения.

### *Требования к Анализу - Геолокация*

На странице авторизованному пользователю должны быть доступны:

а) ввод исходных данных для анализа списка сообществ:

- сформированного вручную;
- импортированного из excel-файла;
- ранее созданного.

б) создание задачи для выполнения.

### *Требования к Анализу - Социальный граф*

а) ввод исходных данных для анализа:

- списка сообществ:
  - сформированного вручную;
  - импортированного из excel-файла;
  - ранее созданного.
- выбор способа построения графа (по ссылкам сообществ или по общим пользователям).

б) создание задачи для выполнения.

## **1.4 Выводы по аналитическому обзору научной и технической документации**

Проведенный аналитический обзор предметной области, в том числе обзор существующих программных продуктов для поиска целевых сообществ

в социальных сетях, позволяет сделать вывод, что существующие веб-сервисы не позволяют находить сообщества, относящиеся к конкретной предметной области, а также отсутствуют инструментарий для анализа сообществ и пользователей социальной сети.

Опираясь на проведенный обзор и анализ предметной области, можно сформулировать цель настоящей работы следующим образом: разработать алгоритмическое и программное обеспечение для повышения эффективности поиска и последующего анализа целевых сообществ в социальных сетях.

## **2 Разработка архитектуры и алгоритмического обеспечения для поиска и анализа целевых сообществ**

### **2.1 Высокоуровневая архитектура программного обеспечения**

Сегодня пользователи Интернета, которые производят специфический контент, используют современные цифровые технологии также как обычные пользователи. Этот факт делает новый вызов перед службами общественного контроля. Отслеживание цифровых следов пользователей Интернета по-прежнему является актуальной проблемой, которая требует комплексного решения и которое может быть использовано для агрегирования, анализа и визуализация. Предлагаемое решение является специализированной поисковой системой, способной найти пользователя или сообщество пользователей по контексту их постов и комментариев, а также анализировать активность на конкретной странице. Целью разработки поисковой системы является предоставление набора инструментов для определения объектов (пользователей или сообществ), которые производят специфический контент в социальных сетях.

Высокоуровневая архитектура системы автоматизированного поиска целевых сообществ в социальных сетях представлена на рисунке 2.1 и описана в публикации [17].

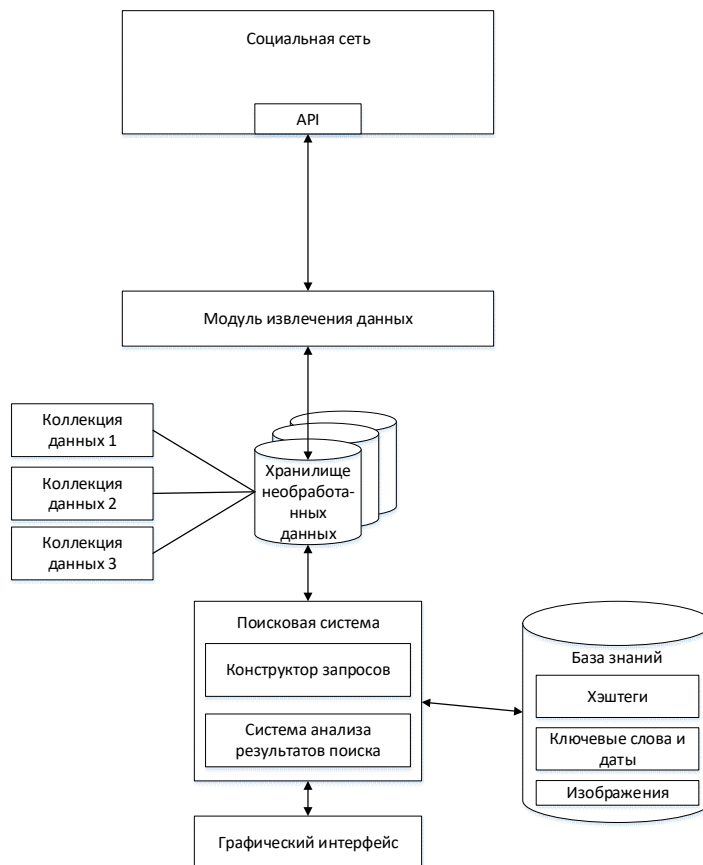


Рисунок 2.1 - Высокоуровневая архитектура системы автоматизированного поиска целевых сообществ в социальных сетях

### 2.1.1 Получение данных из социальной сети

Существует несколько решений для получения интересующей нас информации из социальных сетей. Среди них можно выделить сервисы, которые работают через API (от англ. application programming interface) и методы web scraping.

API – это интерфейс взаимодействия между социальной сетью и сторонними программами, набор готовых классов, функций, процедур, структур и констант, предоставляемых приложением (библиотекой, сервисом) или операционной системой для использования во внешних программных продуктах.

У самой популярной социальной сети на территории СНГ ВКонтакте тоже есть API. Система для разработчиков сторонних сайтов Open API даёт возможность легко авторизовывать пользователей ВКонтакте на сайте. Кроме

того, вы можете получить доступ к информации о друзьях, видеороликах, аудиозаписях, фотографиях и прочих данных ВКонтакте, с согласия пользователей.

В рамках работы с Open API создается специальное приложение, которое позволяет использовать все открытые методы API ВКонтакте. Кроме этого, если уже есть учетная запись ВКонтакте, Open API предоставляет возможность упростить процесс регистрации новых пользователей.

API ВКонтакте — это интерфейс, который с помощью http-запросов к специальному серверу позволяет получать информацию из базы данных ВКонтакте. Нам не нужно знать из каких таблиц и полей каких типов она состоит — достаточно того, что API-запрос это знает. Синтаксис запросов и тип возвращаемых ими данных строго определены.

Несмотря на все достоинства API, существует недостаток - поскольку социальные сети стараются сохранять приватность пользователей, а также некоторые функции слишком сильно нагружают серверную часть приложения, социальная сеть не может отдавать все данные, которые видны в интерфейсе пользователям [18].

### **2.1.2 Хранилище данных**

Агрегированные данные из социальных сетей, извлеченные с использованием API, относятся к неструктурированным или полуструктурированным данным, которые вызывают набор требований для хранилища данных. Неструктурированные данные представленные в виде текста могут включать в себя другие фрагменты данных для будущего анализа. В то же время полуструктурированные данные представлены в некоторых форматы, которые не требуют конкретной схемы (например, XML и JSON). Из-за неопределенной сложности анализируемых текстовых данных, формат файлов JSON представляется наиболее подходящим.

Хранилище необработанных данных является ключевым элементом системы который хранит необработанные данные из социальных сетей для будущего анализ.

Поисковая система рассматривает гибкую модель данных, потому что пользователи могут выполнять свою учетную запись в социальных сетях частично по собственному желанию. Кроме того, во время анализа, количество атрибутов может быть изменено.

Стоит отметить, что основная проблема использования систем управления базами данных это нормализация больших объемов данных. Нормальная форма, в реляционная модель данных, рассматривает представления данных с минимальный уровень лишних данных; однако из-за структурированной формы, этот способ представления данных является трудоемкой задачей. В этом случае, процесс нормализации также приводит к огромному росту общего числа таблиц, поэтому слишком сложно определить пересечения анализируемых объектов (групп или пользователей).

### **2.1.3 База знаний**

Поисковая система проверяет необработанные данные, используя базу знаний, которая содержит predetermined наборы ключевых слов, хэштеги и картинки. Основной задачей базы знаний является содержание классификации, для реализации которой статистические корреляции используются для построения правил размещения объектов в конкретные классы контента. Вопрос классификации является задачей распознавания, где для контрольного образца система присваивает новый объект для определенной категории. Для создания учебного набора и построение модели классификатора методы на основе машинного обучения, как правило, требует формирования словаря с помеченными ключевыми словами. Положительным примером считаются маркированные ключевые слова. Остальные слова считаются отрицательным образцом. Актуальность каждого слова определяется такими параметрами, как, длина слова, положения слова в заголовке, положение в абзаце, в списке

литературы. Далее, вычисляется частота каждого слова из группы ключевых слов и устанавливается порог. Для нового документа производится расчет актуальности ключевых слов и частоты их появления.

Следует отметить, что при соответствующих изменениях в базе знаний, возможен поиск целевых сообществ различных категорий. Наполнение базы знаний информацией по целевым сообществам разных категорий позволит изучить характер и силу их взаимосвязей и взаимного влияния.

## **2.2 Алгоритмическое обеспечение для поиска целевых сообществ**

### **2.2.1 Алгоритм поиска по ключевым словам**

Прямой поиск сообществ основан на использовании стандартного метода социальной сети Вконтакте *groups.search*. Главным параметром данного метода является текст поискового запроса, в который мы, используя цикл, поочерёдно подаём полученные ранее ключевые слова рис. 2.1.1.

Принцип работы заключается в следующем:

- 1) Подача сформированного списка ключевых слов в стандартный метод социальной сети Вконтакте *groups.search*.
- 2) Получение найденных сообществ, соответствующих ключевым словам.
- 3) Вывод результата.

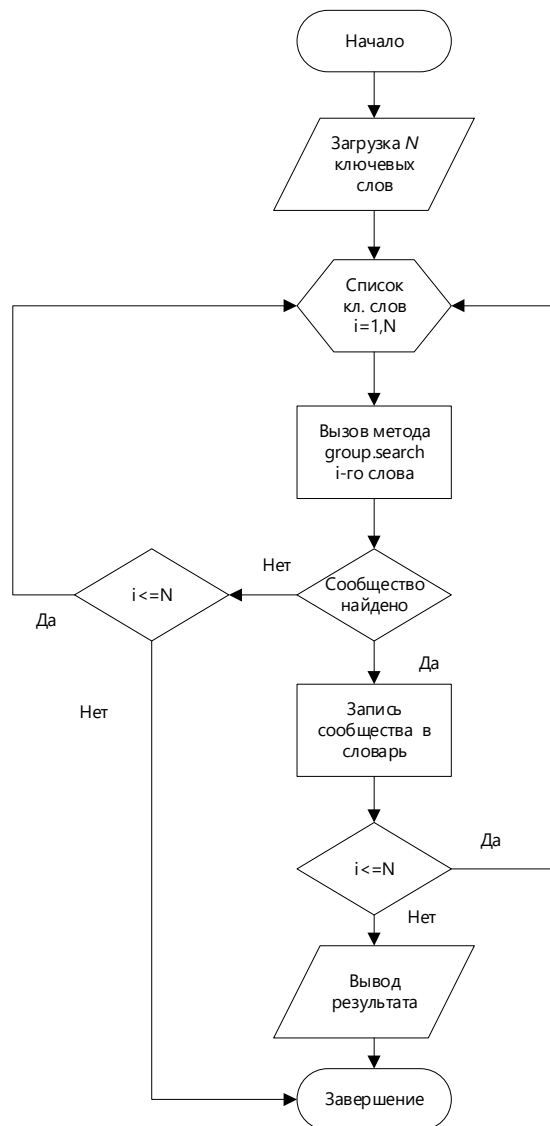


Рисунок 2.1.1 – Алгоритм прямого поиска сообществ

### 2.2.2 Алгоритм поиска смежных сообществ

После нахождения целевого сообщества существует высокая вероятность того, что целевая аудитория найденного сообщества находится в сообществах с подобными интересами. Для реализации, проверки и использования данной гипотезы был разработан алгоритм поиска смежных сообществ рис.2.2.1.

Принцип работы заключается в следующем:

- 1) Получение ID сообщества при помощи стандартного метода социальной сети Вконтакте *groups.getById* по известному *screen\_name* целевого сообщества .

- 2) Получение списка пользователей, подписанных на данное сообщество при помощи метода *groups.getMembers*.
- 3) Получение списка сообществ каждого пользователя при помощи метода *groups.get* и запись их в словарь. Результатом данного шага будет словарь, в котором будут записаны названия сообществ и сколько раз они встретилось у пользователей. Данная статистика будет собрана только для открытых профилей пользователей.
- 4) Сортировка списка смежных сообществ пользователей исходного целевого сообщества.
- 5) Вывод результата.





Рисунок 2.2.1 – Алгоритм поиска смежных сообществ

## 2.3 Алгоритмическое обеспечения анализа целевых сообществ

### 2.3.1 Алгоритмы формирования «профиля» группы

Для анализа целевого сообщества разработаны два алгоритма - алгоритм формирования возрастной, гендерной характеристики и геолокации и алгоритм построения социального графа с вычислением центральных вершин рисунок 2.3.1 и рисунок 2.3.2.

Принцип работы алгоритма формирования характеристик заключается в следующем:

- 1) Запрос информации о каждом пользователе сообщества и запись её в список.
- 2) Группировка по возрасту, расчёт медианы и среднего возраста пользователей сообщества.
- 3) Группировка по полу и расчёт соотношения.
- 4) Группировка по геолокации и построение гистограммы.
- 5) Вывод результата.

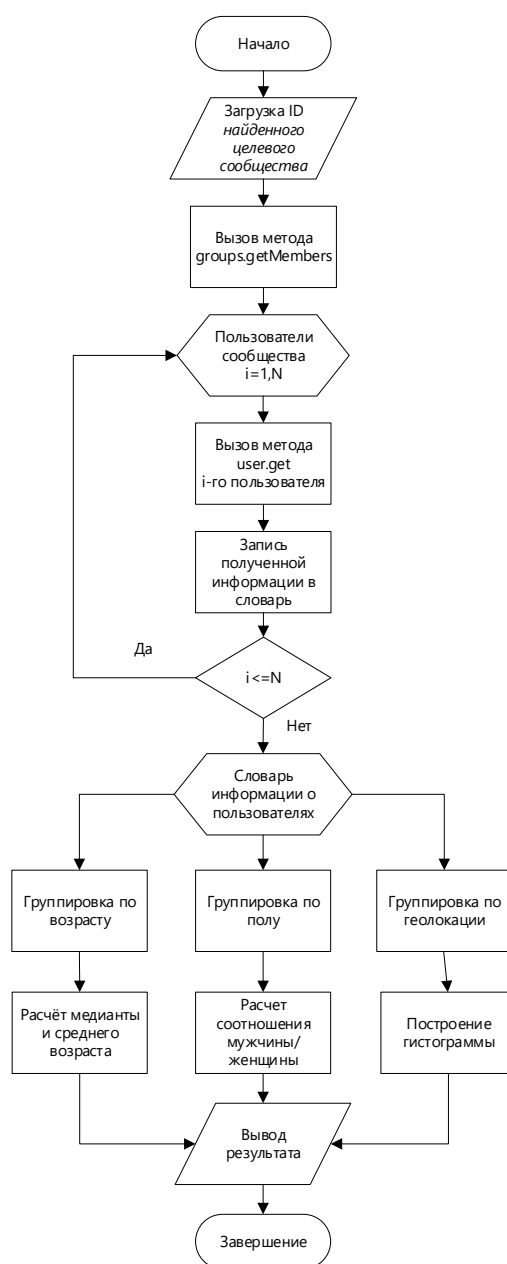


Рисунок 2.3.1 – Алгоритм формирования характеристик сообщества

Принцип работы алгоритма построения социального графа заключается в следующем:

- 1) Запрос информации о каждом пользователе сообщества и запись её в список.
- 2) Запрос информации о друзьях каждого пользователя.
- 3) Построение матрицы смежности сообщества.
- 4) Построение социального графа сообщества по матрице смежности.
- 5) Вычисление степени центральности, близость к центру (closeness centrality), степень посредничества (betweenness centrality), центральность собственного вектора (eigenvector centrality).
- 6) Вывод результата.

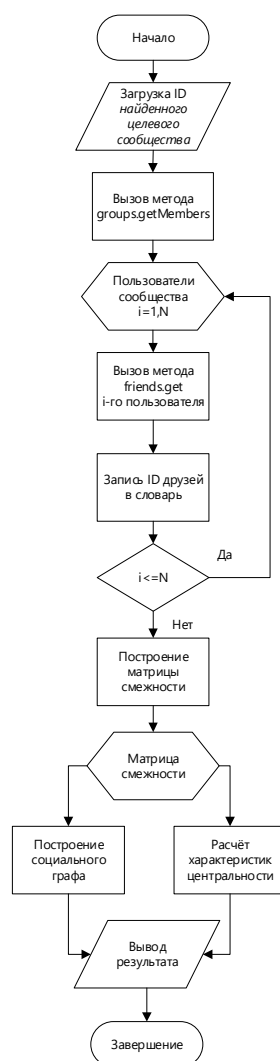


Рисунок 2.3.2 – Алгоритм построения социального графа

### 2.3.2 ККА или алгоритм анализа календарной активности сообщества

Для поиска целевых сообществ в социальной сети Вконтакте был спроектирован и реализован алгоритм календарно-корреляционного анализа активности сообществ. Принцип работы алгоритма приведён на рис.2.3.1.

По предварительно проведённым исследованиям целевых сообществ были получены два файла- ключевые слова, ключевые даты.

Принцип работы заключается в следующем:

- 1) Запрос записей анализируемого сообщества из социальной сети Вконтакте, используя стандартный метод библиотеки Вконтакте *wall.get*, за временной интервал – 1 календарный год.
- 2) Считывание файла с ключевыми датами. Поиск записей для каждой ключевой даты и анализ на количество встречаемых ключевых слов в данной ключевой дате. Запись результата в отдельный файл.
- 3) Анализ полученных записей сообщества за один календарный год на количество встречаемых ключевых слов. Результат – сколько каждое ключевое слово встречается в записях за один календарный год, общее количество встречаемых ключевых слов за один календарный год. Запись результата в файл.
- 4) Выставление экспертного порога по количеству найденных записей, соответствующих ключевым датам, и общему количеству найденных ключевых слов в записях сообщества за один календарный год.
- 5) Анализ результатов на превышение порога.
- 6) Вывод результата.

Был написан и опробован дополнительный алгоритм анализа комментариев к постам. Результат показал его неэффективность, т.к. алгоритму необходимо значительно большее количество запросов к базе данных социальной сети, а результат незначителен.



Рисунок 2.3.3 – Алгоритм ККА

### 2.3.3 Визуализация геолокационных данных

Для построения геолокации целевых сообществ на карте был спроектирован и реализован алгоритм. Принцип работы алгоритма приведён на рис.2.3.4.

Принцип работы заключается в следующем:

- 1) Запрос информации о каждом пользователе сообщества и запись её список.

- 2) Запрос географических координат для полученных местоположений пользователей.
- 3) Нанесение точек на карту.
- 4) Отрисовка карты.
- 5) Вывод результата.

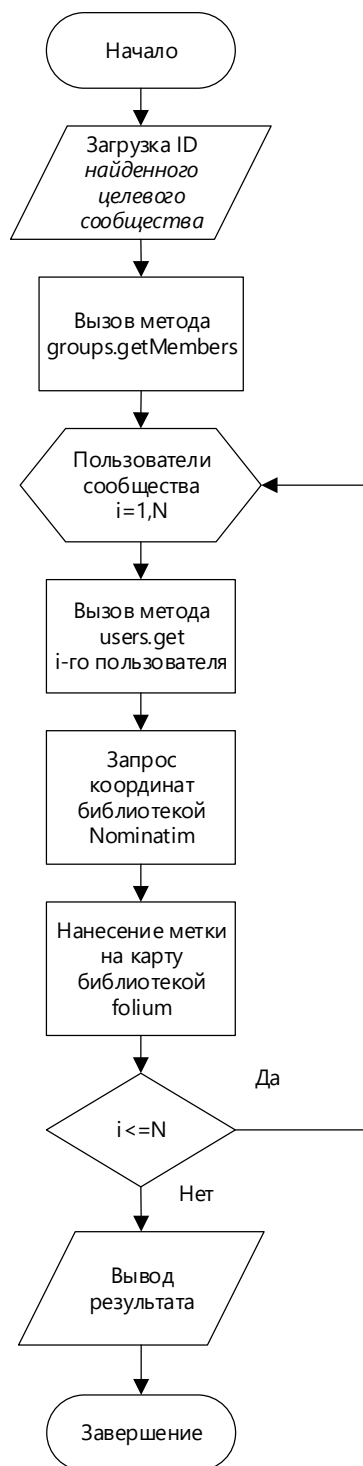


Рисунок 2.3.4 – Алгоритм визуализации геолокационных данных

## **3 Реализация программного обеспечения для поиска и анализа целевых сообществ**

### **3.1 Краткое описание используемых технологий**

#### **3.1.1 Язык программирования Python 3**

Главное преимущество языка Python заключается в читаемости, ясности и более высоком качестве, отличающими его от других инструментов в мире языков сценариев. Программный код на языке Python читается легче, а значит, его использование и обслуживание гораздо проще, чем использование на других языках.

В составе Python поставляется стандартная библиотека с большим числом собранных и переносимых функциональных возможностей. Эта библиотека предоставляет много возможностей, необходимых в прикладных задачах, начиная от поиска текста по документу и заканчивая сетевыми функциями. Выбор языка программирования в первую очередь осуществлялся из соображений минимизации времени разработки основных программных функций.

#### **3.1.2 Фреймворк Django**

Django (Джанго) — свободный фреймворк для веб-приложений на языке Python, использующий шаблон проектирования MVC [19].

Сайт на Django строится из одного или нескольких приложений, которые рекомендуется делать отчуждаемыми и подключаемыми. Это одно из существенных архитектурных отличий этого фреймворка от некоторых других (например, Ruby on Rails). Один из основных принципов фреймворка — DRY (англ. Don't repeat yourself). Также, в отличие от других фреймворков, обработчики URL в Django конфигурируются явно при помощи регулярных выражений. Для работы с базой данных Django использует собственный ORM, в котором модель данных описывается классами Python, и по ней генерируется схема базы данных.

### 3.1.3 HTML

HTML (от англ. HyperText Markup Language — «язык гипертекстовой разметки») — стандартизированный язык разметки документов во Всемирной паутине [20]. Большинство веб-страниц содержат описание разметки на языке HTML (или XHTML). Язык HTML интерпретируется браузерами; полученный в результате интерпретации форматированный текст отображается на экране монитора компьютера или мобильного устройства.

Язык HTML до 5-й версии определялся как приложение SGML (стандартного обобщённого языка разметки по стандарту ISO 8879). Спецификации HTML5 формулируются в терминах DOM (объектной модели документа).

Язык XHTML является более строгим вариантом HTML, он следует синтаксису XML и является приложением языка XML в области разметки гипертекста.

Во всемирной паутине HTML-страницы, как правило, передаются браузерам от сервера по протоколам HTTP или HTTPS, в виде простого текста или с использованием шифрования.

### 3.1.4 CSS

CSS (англ. Cascading Style Sheets — каскадные таблицы стилей) — формальный язык описания внешнего вида документа, написанного с использованием языка разметки [21].

Преимущественно используется как средство описания, оформления внешнего вида веб-страниц, написанных с помощью языков разметки HTML и XHTML, но может также применяться к любым XML-документам, например, к SVG или XUL.

CSS используется создателями веб-страниц для задания цветов, шрифтов, стилей, расположения отдельных блоков и других аспектов представления внешнего вида этих веб-страниц. Основной целью разработки CSS являлось



отделение описания логической структуры веб-страницы (которое производится с помощью HTML или других языков разметки) от описания внешнего вида этой веб-страницы (которое теперь производится с помощью формального языка CSS). Такое разделение может увеличить доступность документа, предоставить большую гибкость и возможность управления его представлением, а также уменьшить сложность и повторяемость в структурном содержимом.

Кроме того, CSS позволяет представлять один и тот же документ в различных стилях или методах вывода, таких как экранное представление, печатное представление, чтение голосом (специальным голосовым браузером или программой чтения с экрана), или при выводе устройствами, использующими шрифт Брайля.

### **3.1.5 BootStrap**

Bootstrap — свободный набор инструментов для создания сайтов и веб-приложений [22]. Включает в себя HTML- и CSS-шаблоны оформления для типографики, веб-форм, кнопок, меток, блоков навигации и прочих компонентов веб-интерфейса, включая JavaScript-расширения.

Bootstrap использует современные наработки в области CSS и HTML, поэтому необходимо быть внимательным при поддержке старых браузеров.

Основные инструменты Bootstrap:

Сетки — заранее заданные размеры колонок, которые можно сразу же использовать, например, ширина колонки 140 px относится к классу `.span2` (`.col-md-2` в третьей версии фреймворка), который можно использовать в CSS-описании документа.

Шаблоны — фиксированный или резиновый шаблон документа.

Типографика — описания шрифтов, определение некоторых классов для шрифтов, таких как код, цитаты и т. п.

Медиа — предоставляет некоторое управление изображениями и видео.

Таблицы — средства оформления таблиц, вплоть до добавления функциональности сортировки.

Формы — классы для оформления форм и некоторых событий, происходящих с ними.

Навигация — классы оформления для панелей, вкладок, перехода по страницам, меню и панели инструментов.

Алерты — оформление диалоговых окон, подсказок и всплывающих окон.

### **3.1.6 База данных Mongo**

Нереляционная документно-ориентированная база данных MongoDB представляет собой удобное, легко развертываемое решение для проектов, в которых предполагается изменения схемы данных с течением времени и наличие пустых полей (иными словами, хранилище не зависит от схемы данных).

Согласно CAP-теореме, MongoDB поддерживает согласованность данных — т.е. на всех вычислительных узлах в любой момент времени находятся одинаковые данные, — а также устойчивость к разделению данных в распределенной системе. В то же время MongoDB, согласно CAP-теореме, не поддерживает свойство устойчивости к падению узлов — в данном случае произойдет отказ системы.

Объемы собираемых метаданных из социальных сетей являются сравнительно небольшими, поэтому MongoDB хорошо подходит для решения задачи хранения подобных метаданных. Данные разбиваются на соответствующие коллекции, в которых каждая запись является отдельным JSON-подобным документом.

### **3.1.7 Система контроля версий git и облачный репозиторий github.com**

Git — распределённая система управления версиями. Основное отличие Git'а от любой другой аналогичной системы это подход Git'а к работе со своими данными. Концептуально, большинство других систем хранят информацию в виде списка изменений в файлах. Эти системы (CVS, Subversion, Perforce, Bazaar и т.д.) представляют хранимую информацию в виде набора файлов и изменений, сделанных в каждом файле, по времени (обычно это называют контролем версий,

основанным на различиях). Применение git позволяет контролировать изменения, вносимые в код, что особенно удобно при коллективной работе над проектом.

GitHub — крупнейший веб-сервис для хостинга IT-проектов и их совместной разработки. Данный сервис предоставляет хранилище для кода проекта, а также инструменты управления разработкой, такие как:

- Kanban доски;
- баг-трекер;
- база знаний проекта;
- разграничение прав доступа пользователей к разным веткам

репозитория.

Совместное использование Git и GitHub позволяет значительно укорить разработку проекта в команде состоящей из нескольких программистов.

### 3.2 Программная реализация пользовательского интерфейса

Разработка web-приложение велась на языке программирования Python 3 с использованием фреймворка Django.

Приведём пример страницы авторизации разработанного Web-приложения рисунок 3.2.1.

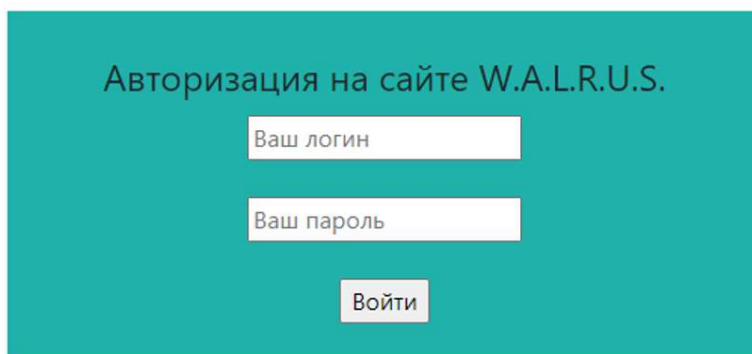


Рисунок 3.2.1 – Страница авторизации

При успешной авторизации и входе в web-приложение попадаем на главную страницу рисунок 3.2.2.

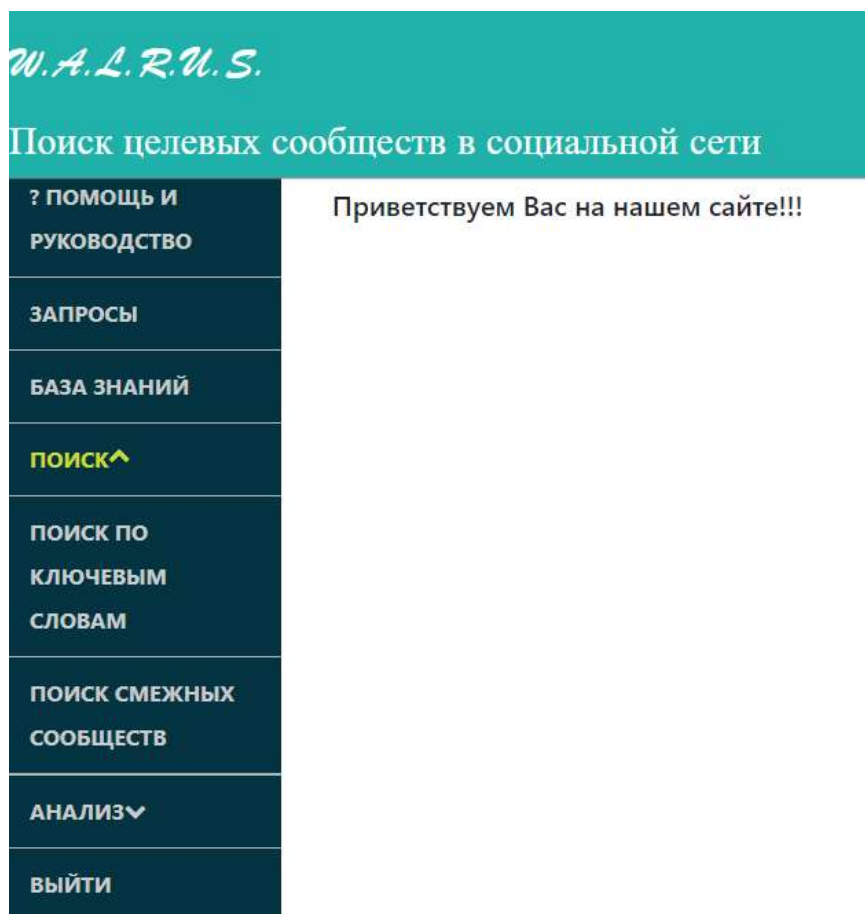


Рисунок 3.2.2 – Главная страница web-приложения

На главной странице находится меню, в котором представлены все функции разработанного приложения.

Раздел меню «Помощь и руководство» содержит справочную информацию о программном обеспечении, описание возможностей программного обеспечения, параметров запросов, проблем и способов их решения.

Раздел меню «Запросы» содержит информацию о всех запросах, сделанных пользователем. По каждому запросу отображается информация когда он был запущен, когда завершён, либо какой процент выполнен и ориентировочное время окончания. По выполненному запросу доступен файл результата запроса.

Раздел меню «База знаний» содержит файл ключевых слов и дат и файл последней версии базы целевых сообществ. Данные файлы доступны по ссылке. Страница «База знаний» приведена на рисунке 3.2.3.

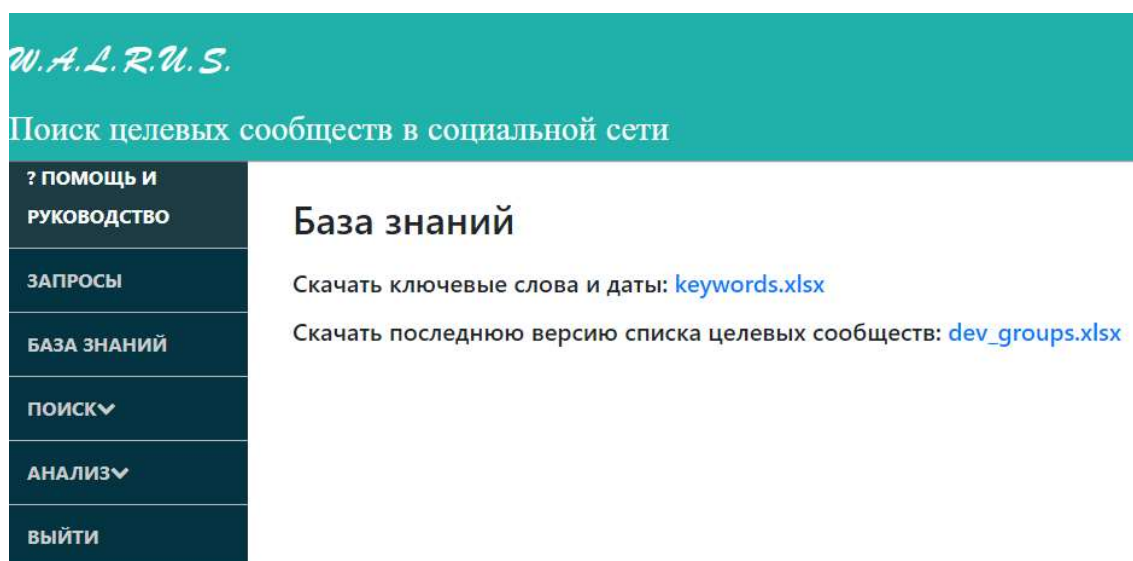


Рисунок 3.2.3 – Страница «База знаний»

Раздел меню «Поиск» содержит две функции – «Поиск по ключевым словам» и «Поиск смежных сообществ». На рисунке 3.2.4 приведена страница «Поиск по ключевым словам», на которой отображены настройки параметров поиска.

W.A.L.R.U.S.

Поиск целевых сообществ в социальной сети

? ПОМОЩЬ И РУКОВОДСТВО

ЗАПРОСЫ

БАЗА ЗНАНИЙ

ПОИСК▼

АНАЛИЗ▼

ВЫЙТИ

### Поиск целевых сообществ по ключевым словам

☐ Загрузить файл ключевых слов и дат  Файл не выбран

☐ Введите ключевое слово

Введите дату активности с   Введите дату активности по

Участников от  Участников до

Рисунок 3.2.4 – Страница «Поиск по ключевым словам»\

Раздел меню «Геолокация» содержит настройки параметров построения геолокации пользователей рисунок 3.2.5. При заполнении необходимых полей и нажатии кнопки «Построить на карте», внизу страницы выводится результат.

W.A.L.R.U.S.

Поиск целевых сообществ в социальной сети

? ПОМОЩЬ И РУКОВОДСТВО

ЗАПРОСЫ

БАЗА ЗНАНИЙ

ПОИСК▼

АНАЛИЗ▼

ВЫЙТИ

### Геолокация пользователей

☐ Загрузить файл целевых сообществ  Файл не выбран

☒ Введите ID сообществ

☐ Введите ID пользователя

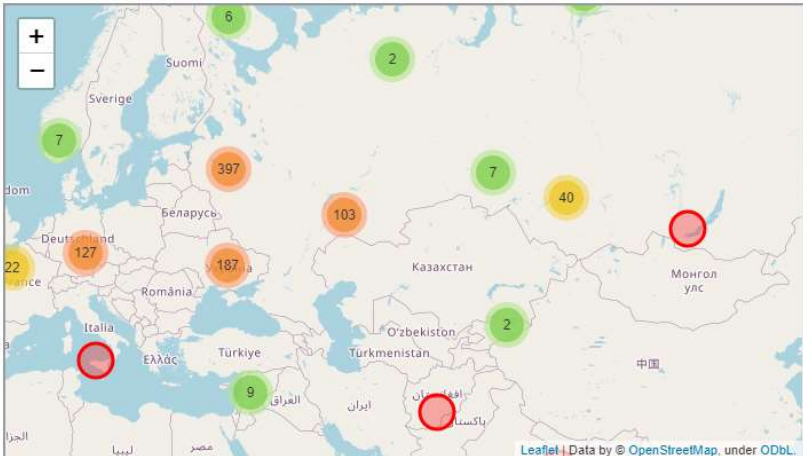


Рисунок 3.2.5 – Страница «Геолокация»

Другие разделы меню аналогичным образом содержать параметры соответствующих функций разработанного приложения.

### **3.3 Апробация программного обеспечения**

Для оценки эффективности работы алгоритма были выполнены тестовые испытания:

В рамках первого испытания осуществлён поиск сообществ по подготовленным формализованным ключевым словам (49 слов). Найдено 6151 сообщество. Сообщества проверены методом ККА, который указал только на 3 сообщества. Все они относятся к целевым сообществам. Что и требовалось проверить (учитывались только сообщества, проявляющие активность в течение календарного года до даты анализа).

При втором испытании был проверен ранее упоминавшийся экспертом сформирован список из 94 сообществ. При тестировании данного списка сообществ обнаружено, что из них активных – 68, заблокировано - 3, закрыто – 1, не найдено – 2. При определении активности принято: если сообщество публикует одну запись в неделю, оно считается активным. Из указанного списка ККА «сработал» на 14 сообществах.

В рамках третьей проверки алгоритма был выполнен тестовый вариант «обратного ККА». Выбраны 39 наиболее популярных сообщества по версии ВКонтакте. Эксперты проверили их вручную. Ни одно из сообществ экспертом не было отнесено к целевым. Апробация показала, что на данных сообществах «чувствительность» алгоритма не сработала. Следовательно, эффективность работы алгоритма подтверждена.

Эффективность работы алгоритма может также быть проиллюстрирована через значения ошибок первого и второго рода. Они могут быть вычислены для второй и третьей проверок, поскольку априорная информация о степени принадлежности к целевым сообществам из 6151 групп, найденных по ключевым словам, отсутствует. В качестве нулевой гипотезы  $H_0$  примем, что анализируемое ВК-сообщество не содержит признаков принадлежности к

целевым сообществам. Соответственно, альтернативная гипотеза –  $H_1$  – ВК-сообщество содержит признаки принадлежности.

В рамках второй проверки гипотеза  $H_0$  была ошибочно принята в 54 случаях, гипотеза  $H_1$ , соответственно, верно принята в 14. Таким образом, вероятность ошибки второго рода равна  $\beta=0,79$  и мощность критерия  $(1-\beta)=0,21$ . Сравнительно низкое значение мощности критерия объясняется неполнотой информации, представленной в базе знаний и описывающей ключевые слова и даты целевых сообществ.

В рамках третьей проверки гипотеза  $H_0$  ни разу не была ошибочно отвергнута и гипотеза  $H_1$  не была ошибочно принята, т.е. ошибок применения критерия не обнаружено.

В общем виде, последовательность автоматизированного поиска целевого сообщества может быть представлена следующей последовательностью:

а) Пользователь-эксперт формирует «базу знаний» (перечень ключевых слов, выражений и дат, в том числе, определяя их взаимосвязь).

б) Пользователь запускает функцию первичного поиска по ключевым словам.

в) Предварительная обработка полученных результатов (удаление закрытых, неактивных, «пустых» сообществ).

г) Методом ККА осуществляется уточнение списка выявленных целевых сообществ.

д) Результаты анализируются пользователем и при необходимости заносятся в базу знаний.

е) Пользователь формирует множество групп для последующего поиска смежных сообществ (сателлитов).

ж) Результаты поиска сообществ-сателлитов анализируются. При необходимости уточняется информация в базе знаний.

Обобщенная схема работы алгоритма представлена на рисунке 3.3.1.



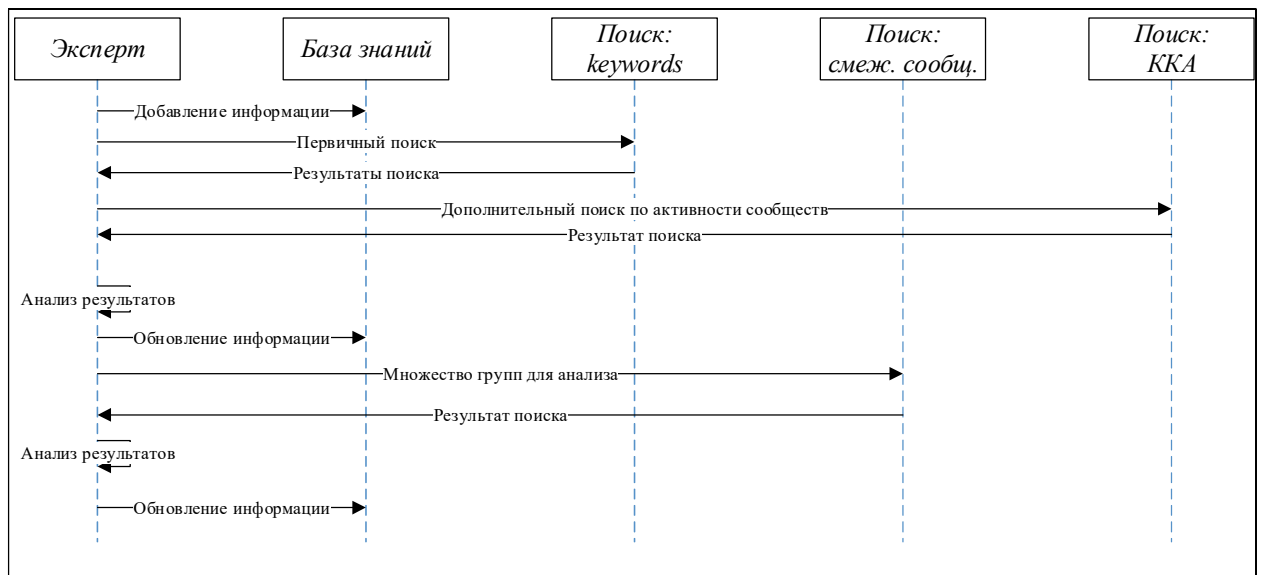


Рисунок 3.3.1 - Диаграмма последовательности обобщенного алгоритма выявления активных целевых сообществ в социальной сети.

### **3.4 Заключение по разделу**

При прохождении преддипломной практики был проведён аналитический обзор предметной области. Это позволило понять, что сделано на данный момент, в каком направлении ведутся исследования и какие проблемы могут возникнуть при разработке алгоритмического и программного обеспечения поиска и анализа целевых сообществ в социальной сети.

Спроектирован, разработан и апробирован алгоритм построения геолокации пользователей и сообществ социальной сети на карте.

Спроектирован и разработан интерфейс программного обеспечения поиска и анализа целевых сообществ в социальной сети.

Результаты работы применимы в обеспечении информационной поддержки при проведении социологических и политологических исследований, в т.ч. при изучении вопросов информационной безопасности.

А так же:

- Изучение и моделирование процессов информационной диффузии в социальных сетях.
- Оценка реакции аудитории студентов и абитуриентов на новости и события.
- Поддержка процессов продвижения актуальной информационной повестки.
- Поиск целевых сообществ абитуриентов для повышения эффективности приемной кампании.

Данные работы выполнялись в рамках ВИУ-УОД-145 2019 и гранта РФФИ 19-011-31535.

Получен Акт внедрения.

## **4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

Данная магистерская диссертация направлена на создание алгоритмического и программного обеспечения (ПО), которое позволит осуществлять поиск целевых сообществ в социальной сети для исследований в области общественно-политических наук, либо с маркетинговой целью.

Целью данного раздела является проведение комплексного описания и анализа финансово-экономических аспектов выполненной работы. А именно, необходимо осуществить планирование процесса управления научно-техническим исследованием (НТИ), рассчитать полные денежные затраты на проект, на основе которых впоследствии дать оценку экономической эффективности и целесообразности осуществления работы.

### **4.1 Организация и планирование работ**

При организации процесса реализации магистерской работы необходимо рационально планировать занятость каждого из участников процесса и сроки проведения отдельных работ. Для этого был составлен полный перечень проводимых работ. Для каждой работы определены исполнители и их нагрузка.

Исполнителями являются следующие участники процесса написания магистерской диссертации:

- инженер – автор магистерской диссертации (И);
- научный руководитель (НР).

План работ и распределение нагрузки между исполнителями представлены в таблице 4.1.

Таблица 4.1 – Перечень работ и продолжительность их выполнения

№ этапа	Этапы работы	Исполнители	Загрузка исполнителей
1	Постановка целей и задач	НР	НР – 100%
2	Анализ предметной области	НР, И	НР – 50% И – 50%
3	Составление и утверждение ТЗ	НР, И	НР – 30% И – 100%
4	Разработка календарного плана	НР, И	НР – 100% И – 10%
5	Высокоуровневая архитектура программного обеспечения	НР, И	НР – 80% И – 100%
6	Алгоритмическое обеспечение для поиска целевых сообществ	И	И – 100%
7	Алгоритмическое обеспечения анализа целевых сообществ	И	И – 100%
8	Проектирование пользовательского интерфейса	И	И – 100%
9	Апробация программного обеспечения	И	И – 100%
10	Краткое описание используемых технологий	И	И – 100%
11	Подготовка отчетной документации	И	И – 100%
12	Получение акта внедрения результатов магистерской диссертации	НР	НР – 100%
13	Защита диссертации	И	И – 100%

В результате анализа перечня работ определено 13 этапов, которые распределены на два исполнителя: инженера (автора работы) и научного руководителя.

#### 4.1.1 Продолжительность этапов работ

После выявления всех этапов научно-исследовательской работы магистранта необходимо рассчитать их продолжительность. Для формирования календарного плана выполнения работ был использован опытно-статистический метод, который реализуется двумя способами:

- аналоговый;
- экспертный.

Ожидаемые значения продолжительности работ тоже были рассчитаны с использованием экспертного способа, реализуемого при помощи следующей формулы:

$$t_{ож} = \frac{3 \cdot t_{\min} + 2 \cdot t_{\max}}{5},$$

где  $t_{\min}$  – минимальная продолжительность работы, дн.;

$t_{\max}$  – максимальная продолжительность работы, дн.;

Исполнители работ были определены в предыдущем пункте, это инженер и научный руководитель.

Для построения графика работ необходимо рассчитать продолжительность этапов в рабочих днях (по формуле 4.1), а затем перевести полученные результаты в календарные дни, используя формулу 4.2. Формулы расчета приведены ниже.

Расчет продолжительности выполнения каждого этапа в рабочих днях ( $T_{РД}$ ) ведется по формуле:

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д}, \quad (4.1)$$

где  $t_{ож}$  – продолжительность работы, дн.;

$K_{ВН}$  – коэффициент выполнения работ, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей, в частности ( $K_{ВН} = 1$ );

$K_{Д}$  – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ( $K_{Д} = 1,2$ ).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{РД} \cdot T_{К}, \quad (4.2)$$

где  $T_{КД}$  – продолжительность выполнения этапа в календарных днях;

$T_{К}$  – коэффициент календарности, позволяющий перейти от длительности работ в рабочих днях к их аналогам в календарных днях, и рассчитываемый по формуле:

$$T_{К} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}} = \frac{366}{366 - 52 - 14} = 1,22,$$

где  $T_{КАЛ}$  – календарные дни ( $T_{КАЛ} = 366$ );

$T_{ВД}$  – выходные дни ( $T_{ВД} = 52$ );

$T_{ПД}$  – праздничные дни ( $T_{ПД} = 14$ ).

В таблице 4.2 описаны этапы работы и их трудоемкость по исполнителям, занятым на каждом этапе. По показанию полученных величины трудоемкости этапов по исполнителям построен линейный график осуществления проекта, который представлен в таблице 4.2.

Таблица 4.2 – Трудозатраты на выполнение диссертации

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.- дн			
					$T_{РД}$		$T_{КД}$	
		$t_{min}$	$t_{max}$	$t_{ож}$	НР	И	НР	И
1	2	3	4	5	6	7	8	9
Постановка целей и задач	НР	2	4	2,8	3,4	–	4,1	–
Анализ предметной области	НР, И	10	14	11,6	7	7	8,5	8,5
Составление и утверждение ТЗ	НР, И	1	2	1,4	0,5	1,7	0,6	2,1
Разработка календарного плана	НР, И	1	2	1,4	1,7	0,2	2,1	0,2
Высокоуровневая архитектура программного обеспечения	НР, И	2	4	2,8	2,7	3,4	3,3	4,1
Алгоритмическое обеспечение для поиска целевых сообществ	И	20	30	24	–	29	–	35
Алгоритмическое обеспечения анализа целевых сообществ	И	20	30	24	–	29	–	35
Проектирование пользовательского интерфейса	И	40	60	48	2	57,6	2,4	70
Апробация программного обеспечения	И	15	30	21	–	25,2	–	31
Краткое описание используемых технологий	И	4	6	4,8	–	5,8	–	7,1
Подготовка отчетной документации	И	5	10	7	–	8,4	–	10
Получение акта внедрения результатов магистерской диссертации	НР	10	20	14	16,8	–	20,5	–
Защита диссертации	И	1	2	1,4	–	1,7	–	2,1

Итого:				164,2	34,1	169	41,5	205,1
--------	--	--	--	-------	------	-----	------	-------

На основе полученных в таблице данных строится диаграмма Ганта, которая наглядно отображает календарный план-график всех работ, а также показывает нагрузку на исполнителей работ на каждом этапе. Результаты представлены на рисунках 4.1–4.2.

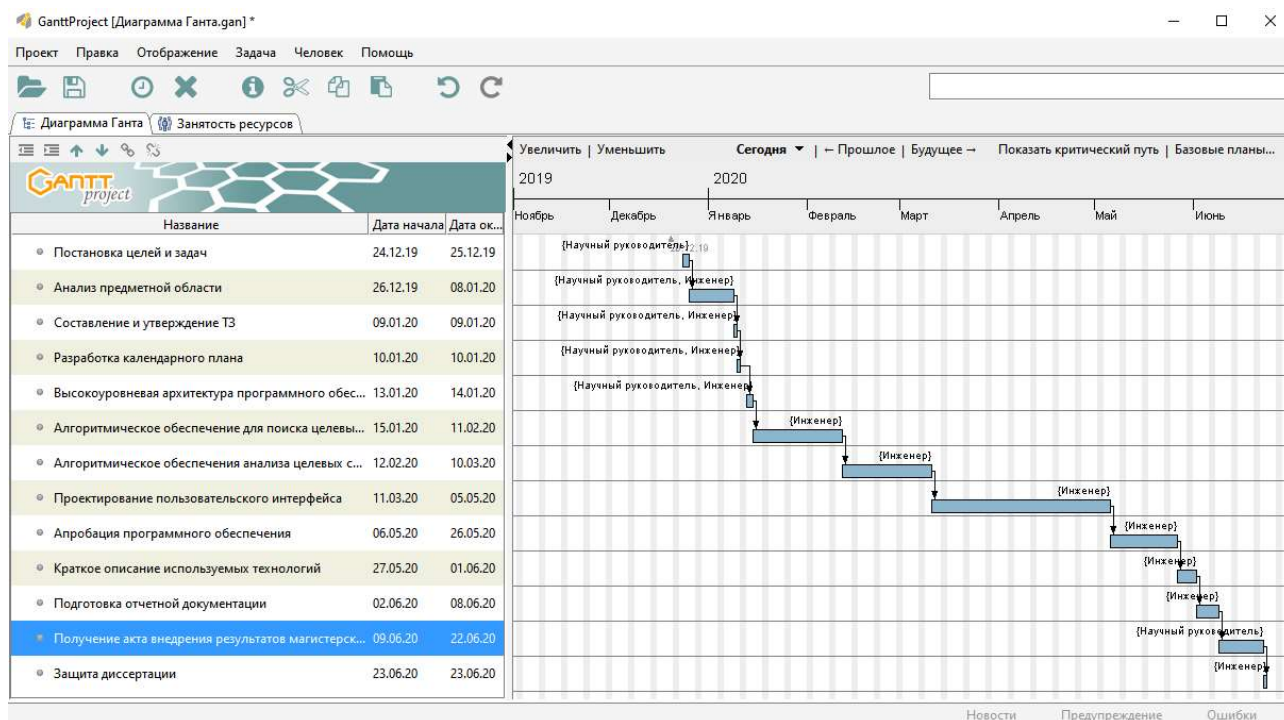


Рисунок 4.1 – График проведения работ

Из диаграммы, приведенной на рис.4.1, видно, что наиболее длительными этапами являются проектирование и реализация интерфейса, разработка алгоритмического обеспечения и апробация программного обеспечения, т.е. непосредственная разработка программного обеспечения информационной системы.

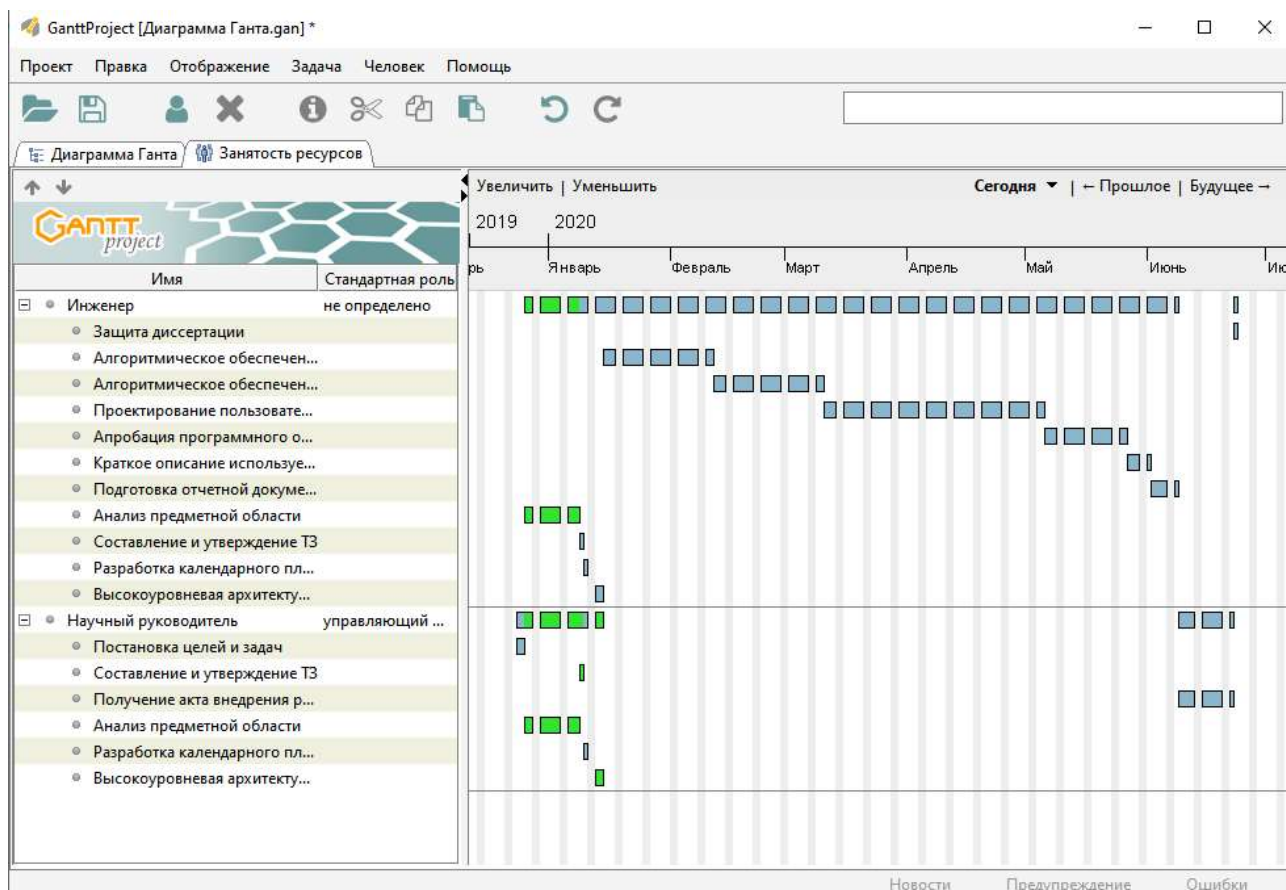


Рисунок 4.2 – Диаграмма занятости ресурсов

Из диаграммы, приведенной на рис.4.2, где зеленым цветом отмечена частичная занятость ресурса, а голубым – полная, видно, что научный руководитель принимает активное участие в работе на ее начальных этапах, курируя работу студента и подсказывая верные направления. При этом инженер на начальных этапах занят частично, а затем, исследовав предметную область и сформировав ТЗ и рабочий план, приступает к выполнению работ с полной занятостью вплоть до защиты работы.

#### 4.2 Расчет сметы затрат на выполнение НИР

Расчет сметы затрат на выполнение НИР необходим для формирования бюджета проекта и включает расчет всех расходов, необходимых для реализации комплекса работ, составляющих содержание данной разработки. В расчет ведется по следующим статьям затрат:

- материалы;



- заработная плата;
- социальный налог;
- расходы на электроэнергию (без освещения);
- амортизационные отчисления;
- прочие (накладные расходы) расходы.

Совокупность перечисленных статей формируют затраты на выполнение НИР.

#### 4.2.1 Расчет затрат на материалы

Статья затрат на материалы включает в себя затраты на приобретение всех видов материалов, необходимых для выполнения НИР по данной теме. Расчет стоимости материальных затрат производится по действующим ценникам или договорам поставки.

Приблизительно они оцениваются в процентах к отпускной цене закупаемых материалов, как правило, это 5%. Расчет затрат на материалы приведен в таблице 4.3.

Таблица 4.3 – Расчёт затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
Бумага А4	2	200 листов	400
<b>Итого:</b>			<b>400</b>

С учетом описанных сопутствующих затрат, составляющих 5% от отпускной цены материалов, расходы на материалы равны  $C_{\text{мат}} = 400 * 1,05 = 420$  руб.

#### 4.2.2 Расчет заработной платы

Данная статья является основной статьей расходов на выполнение НИР. Над НИР работают два исполнителя: научный руководитель и инженер. Среднедневная тарифная заработная плата ( $ЗП_{\text{дн-т}}$ ) в расчете на один рабочий день вычисляется по формуле:

$$ЗП_{\text{дн-т}} = \text{Оклад (месячный)} / 21 ,$$

где учитывается количество рабочих дней в году  $\approx 247$  и, следовательно, в месяце в среднем 21 рабочий день (при пятидневной рабочей неделе).

Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 5.2 (НР – 34,1 = 35, И – 169 = 169 рабочих дней). При этом рабочая неделя пятидневная как для научного руководителя, так и для инженера.

Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется, соответственно, следующий ряд коэффициентов:  $K_{ПР} = 1,1$ ;  $K_{доп.ЗП} = 1,188$ ;  $K_p = 1,3$  для Томской области. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в НИР, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент  $K_{и} = 1,59$ .

Оклады исполнителей брались из информационной таблице, представленной в методических указаниях по финансовому менеджменту для магистрантов.

Расчеты основной заработной платы работников, непосредственно занятых выполнением проекта, с учетом премий, доплат и районного коэффициента представлены в таблице 4.4 и велись по следующей формуле:

$$ЗП = ЗП_{дн-т} * \text{Кол-во раб. дней} * K_{и}.$$

Таблица 4.4 – Расчет основной заработной платы

Исполнитель	Оклад, руб./мес.	Среднедневная ставка (ЗП <sub>дн-т</sub> ), руб./раб.день	Затраты времени, раб. дни	Коэффициент (K <sub>и</sub> )	Фонд з/п, руб.
НР	33664	1603	35	1,59	89207
И	9489	452	169		121457
Итого:					210664

*Для 5-дневной раб. недели коэф-т = 1,59 – см. МУ в (в старых – 1,62)*

Итого суммарные затраты на заработную плату всем участникам НИР с учетом стандартных окладов исполнителей в зависимости от занимаемой должности и ученого звания составляют  $C_{зп} = 210\,664$  рублей.

#### 4.2.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, являются обязательной статьей расходов и составляют 30% от полной заработной платы по НИР. Следовательно, затраты на ЕСН вычисляются по следующей формуле:

$$C_{\text{соц.}} = C_{\text{зп}} * 0,3 ,$$

где  $C_{\text{зп}}$  – суммарные затраты на заработную плату по проекту.

Итоговые затраты на социальный налог равны  $C_{\text{соц.}} = 210\,664 \text{ руб.} * 0,3 = 63\,199 \text{ руб.}$

#### 4.2.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{\text{эл.об.}} = P_{\text{об}} \cdot t_{\text{об}} \cdot Ц_{\text{э}},$$

где  $P_{\text{об}}$  – мощность, потребляемая оборудованием, кВт;

$t_{\text{об}}$  – время работы оборудования, час;

$Ц_{\text{э}}$  – тариф на 1кВт·час. Текущий тариф Для ТПУ – 6,59

При этом время работы оборудования рассчитывается на основе данных таблицы 4.2 для инженера ( $T_{\text{рд}} = 169$  дней) из расчета, что продолжительность рабочего дня равна 8 часов:

$$t_{\text{об}} = T_{\text{рд}} * K_t,$$

где  $K_t \leq 1$  – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к ТРД. Время проводимое за персональным компьютером равно 2/3 от общего времени работы над проектом.

В свою очередь мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{ОБ}} = P_{\text{НОМ.}} * K_C$$

где  $P_{\text{НОМ.}}$  – номинальная мощность оборудования, кВт (показатели для ПК – 0,5 кВт, для принтера – 0,1 кВт);

$K_C \leq 1$  – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Примем значение  $K_C = 1$ .

Расчет затрат представлен в таблице 4.5 ниже.

Таблица 4.5 – Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования <i>t</i> <sub>об</sub> , час	Потребляемая мощность <i>P</i> <sub>об</sub> , кВт	Тариф на 1кВт·час ЦЭ, руб	Затраты Эоб, руб.
Персональный компьютер	169дней * 8 часов * 2/3 = 901	0,5	6,59	2969
Лазерный принтер	10	0,1		6,59
Итого:				2 976

Итого затраты на электроэнергию за вычетом освещения составляют  $C_{\text{эл}} = 2\,976$  рубль, где основными затратами является оплата электроэнергии, потребляемая персональным компьютером, на котором ведется основная работа по НИР.

#### 4.2.5 Расчет амортизационных расходов

Данная статья предназначена для учета амортизации используемого оборудования за время выполнения НИР. Для расчета амортизационных затрат используется следующая формула:

$$C_{\text{АМ}} = \frac{N_A * C_{\text{ОБ}} * t_{\text{рф}} * n}{F_D},$$

где  $N_A$  – годовая норма амортизации единицы оборудования;

$C_{\text{ОБ}}$  – балансовая стоимость единицы оборудования, руб;

$F_D$  – действительный годовой фонд времени работы соответствующего оборудования, час;

$t_{\text{рф}}$  – фактическое время работы оборудования в ходе выполнения проекта, час;

$n$  – число задействованных однотипных единиц оборудования.

$N_A$  для персонального компьютера составляет обратную величину от срока амортизации равного 2,5 года, следовательно,  $N_A(\text{ПК}) = 1 / 2,5 = 0,4$ . Номинальная стоимость одного ПК составляет 40 тысяч рублей. Для расчета  $F_d(\text{ПК})$  будем учитывать, что количество рабочих дней в 2020 году при пятидневной рабочей неделе равно 247 дней, рабочий день длится 8 часов. Таким образом  $F_d(\text{ПК}) = 247 * 8 = 1976$  часов. Однако ПК занят лишь  $2/3$  от проектного времени, затрачиваемого инженером ( $T_{рд}$  из таблицы 4.2), следовательно,  $t_{рф} = 169 \text{ дней} * 8 \text{ часов} * 2/3 = 901 \text{ часов}$ .

Таким образом, амортизация, начисленная на ПК, равна:

$$C_{AM}(\text{ПК}) = \frac{0,4 * 40\,000 \text{ руб} * 901 \text{ часов} * 1 \text{ ед}}{1976 \text{ часов}} = 7\,296 \text{ руб.}$$

Стоимость простого принтера составляет 5 600 руб., его  $F_d = 500 \text{ час.}$ ;  $N_A = 1 / 2 \text{ года} = 0,5$ ;  $t_{рф} = 10 \text{ часов}$  (из табл. 4.5).

$$C_{AM}(\text{ПР}) = \frac{0,5 * 5\,600 \text{ руб} * 10 \text{ часов} * 1 \text{ ед}}{500 \text{ часов}} = 56 \text{ руб.}$$

Итого общие затраты на амортизацию составляют  $C_{AM} = 7\,352 \text{ руб}$ , где большая часть суммы идет на покрытие износа персонального компьютера.

#### 4.2.6 Расчет прочих расходов

Статья прочих расходов включает расходы на выполнение НИР, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов и рассчитываются по следующей формуле:

$$C_{\text{проч.}} = (C_{\text{мат}} + C_{\text{зп}} + C_{\text{соц}} + C_{\text{эл}} + C_{\text{ам}}) \cdot 0,1$$

Таким образом, прочие расходы равны  $C_{\text{проч.}} = (420 + 210\,664 + 63\,199 + 2\,976 + 7\,352) * 0,1 = 28\,461 \text{ руб.}$

#### 4.2.7 Расчет общей себестоимости разработки

Общая стоимость разработки складывается из всех статей расходов, рассчитанных выше. Проведенные расчеты позволяют определить величину бюджета затрат проекта, на которую будут опираться заказчики данной разработки. Смета затрат на разработку представлена в таблице 4.6.

Таблица 4.6 – Расчет бюджета затрат НИИ

Статья затрат	Сумма, руб.
Материалы	420
Основная заработная плата	210 664
Отчисления в социальные фонды	63 199
Расходы на электроэнергию	2 976
Амортизационные отчисления	7 352
Прочие расходы	28 461
<b>Итого:</b>	<b>313 072</b>

Итоговые затраты на разработку составляют  $C = 313\,072$  рублей. Далее на основе рассчитанной себестоимости разработки определим конечную цену всей НИР.

#### 4.2.8 Расчет прибыли

Для расчета прибыли примем, что ее размер составляет 20% от полной себестоимости НИР, так как проект актуален и востребован среди целевой аудитории. Себестоимость проекта равна затратам на его разработку  $C = 313\,072$  рублей. Тогда прибыль составит:  $313\,072 * 20\% = 62\,614$  рублей.

#### 4.2.9 Расчет НДС

Косвенный налог – налог на добавленную стоимость (НДС) составляет 20% от суммы затрат на разработку и прибыли. Таким образом, сумма налога равна  $(313\,072 + 62\,614) * 0,2 = 375\,686 * 20\% = 75\,137$  рублей.

#### 4.2.10 Цена разработки проекта

На основе проведенных расчетов, определим итоговую цену разработки проекта, которая равна сумме полной себестоимости, прибыли и НДС, в нашем случае:  $C_{кр} = 313\,072 + 62\,614 + 75\,137 = 450\,823$  рублей.

Чтобы определить экономическую эффективность НИР, необходимо провести ее оценку.

### 4.3 Оценка экономической эффективности проекта

Актуальным аспектом качества выполненного проекта является экономическая эффективность его реализации, т.е. соотношение обусловленного ей экономического результата (эффекта) и затрат на разработку проекта. Так как

последние являются единовременными, то рассмотрим частный случай задачи оценки экономической эффективности инвестиций с целью получения определенного результата в будущем и определим срок окупаемости затрат на разработку проекта (инвестиций).

#### 4.3.1 Определение срока окупаемости инвестиций (PP – payback period)

При помощи показателя PP определим продолжительность периода, через который инвестиции окупятся благодаря прибыли от реализации продукта. Чем меньше PP, тем эффективнее проект. Использование показателя предполагает установление для него приемлемого значения как меры эффективности инвестиций. Используется формула

$$PP = n_{цj} + \frac{\Delta PP_{чj}}{PP_{чj+1}}, \quad (4.3)$$

где  $n_{цj}$  – целое число лет, при котором накопленная сумма прибыли наиболее близка к величине инвестиций  $I_0$ , но не превосходит ее;

$\Delta PP_{чj}$  – непокрытая часть инвестиций по истечении  $n_{цj}$  лет реализации проекта;

$PP_{чj+1}$  – прибыль за период, следующий за  $n_{цj}$ -м.

В качестве источника прибыли возьмем цену одной реализованной лицензии разработанного программного обеспечения, равной 100 000 рублей.

Так как цена лицензии включает в себя НДС, то для расчета прибыли необходимо этот налог (НДС = 20%) вычесть, т.е.

100 000 руб. (с учетом НДС) – 120%

$PP_{л(без НДС)}$  – 100%

Тогда цена одной лицензии без НДС равна  $C_{л(без НДС)} = 100\,000 * 100\% / 120\% = 83\,333$  руб

Кроме того, для получения значения балансовой прибыли (прибыль до налогообложения) из  $C_{л(без НДС)}$  необходимо вычесть сопутствующие при продаже лицензии расходы, перечень затрат представлен в таблице 4.7.

Таблица 4.7 – Сопутствующие расходы при продаже одной лицензии ПО

Наименование	Цена за 1 шт. (руб)
Установочный диск 1 шт.	500
Комплект документов для покупателя 1 шт.	1 000
Оплата труда дистрибьютеру лицензий за 1 проданную лицензию (в том числе проведения вводного занятия с пользователями ИС)	7 000
<b>Итого:</b>	<b>8 500</b>

Следовательно, балансовая прибыль от продажи одной лицензии равна  $ПР_{л} = 83\,333 - 8\,500 = 74\,833$  рубля.

Также, чтобы пользоваться ПО необходимо каждый год продлевать лицензию. Цена продления лицензии равна 60 000 рублей, при этом сопутствующие затраты отсутствуют, значит, балансовая прибыль от продажи продления равна ее цене за вычетом НДС = 20%. Тогда  $ПР_{л.пд} = Ц_{л.пд}(\text{без ДС}) = 60\,000 * 100\% / 120\% = 50\,000$  рублей.

Так как для расчетов окупаемости проекта необходима чистая прибыль, то из балансовой прибыли необходимо также вычесть налог на прибыль (чистая прибыль рассчитывается как 83,33% от балансовой прибыли).

Таким образом, прибыль за год считается по следующей формуле:

$$\begin{aligned}
 ПР_{чj} &= (ПР_{л} * n1 + ПР_{л.пд} * n2) * 83,33\% = \\
 &= (74\,833 * n1 + 50\,000 * n2) * 83,33\%,
 \end{aligned}$$

где  $ПР_{л}$  – балансовая прибыль с продажи одной лицензии;

$ПР_{л.пд}$  – балансовая прибыль с продажи одного продления лицензии;

$n1$  – количество проданных лицензий;

$n2$  – количество проданных продлений лицензии.

Расчет накопленных денежных поступлений по проекту приведен в таблице 4.8, где каждый год планируется продавать одну лицензию и продлевать ранее купленные.



Таблица 4.8 – Накопленные денежные поступления по проекту

Год	Инвестиции (руб.)	Балансовая прибыль (до вычета налога) (руб.)	Чистая прибыль (руб.)	Накопленный денежный поток (руб.)
0	-450 823	74 833	62 358	-388 465
1		124 833 = 74 833 + 50 000	104 023	-284 442
2		174 833 = 74 833 + 100 000	145 688	-138 754
3		224 833 = 74 833 + 150 000	187 353	48 599
4		274 833 = 74 833 + 200 000	229 018	277 617
5		324 833 = 74 833 + 250 000	270 683	548 300

Рассчитаем РР, используя формулу 4.3. Из приведенных в таблице 4.8 расчетов следует, что 2-й год эксплуатационного периода дает минимум непокрытого остатка (-138 754 руб.) инвестированной суммы 450 823 руб., следовательно,  $n_{цj} = 2$ . Тогда  $\frac{\Delta \text{ПР}_{чj}}{\text{ПР}_{чj+1}} = 138\,754 / 187\,353 = 0,74$ ; следовательно, РР равен 2,74 года.

Полученное значение говорит о том, что при описанном плане продаж затраты на проект окупятся спустя 2 года 7,4 месяца, и дальше такой проект будет приносить только прибыль, которая может значительно вырасти при условии вовлечения новых клиентов для продажи лицензии продукта. К тому же, такая экономическая эффективность напрямую зависит от количества проданных лицензий ПО в первый и последующие годы.

Так как период окупаемости больше года, то при расчете срока окупаемости РР необходимо учесть изменение ценности денег во времени. А также целесообразно рассчитать величину накопленного чистого эффекта по формуле:

$$NPV = \sum_{j=1}^n \text{ПР}_{чj} - I_0, \quad (4.4)$$

где  $n$  – продолжительность в годах периода оценки эффекта или срок реализации проекта примем 5 лет.

При расчетах по формулам 4.3-4.4 вместо величин  $\Delta \text{ПР}_{чj}$  и  $\text{ПР}_{чj+1}$  необходимо использовать их дисконтированные аналоги, получаемые путем

деления  $\Delta PR_{чj}$  и  $PR_{чj+1}$  на  $(1 + i)^j$ , где  $i$  – ставка дисконтирования (целевой уровень годовой доходности инвестируемых средств) с принятым значением  $i = 0,1$ . Следовательно, произведем перерасчет данных таблицы 4.8, но с учетом убывания реальной стоимости результатов в будущие периоды (годы) относительно периода инвестирования. Расчет дисконтированного срока окупаемости представлен в таблице 4.9.

Таблица 4.9 – Расчет дисконтированного срока окупаемости

Год	Инвестиции	Номинальная прибыль	Коэффициент дисконтирования $1/(1+0,1)^i$	Дисконтированная прибыль	Накопленный денежный поток
0	-450 823	62 358	1	62 358	-388 465
1		104 023	0,9091	94 568	-293 897
2		145 688	0,8264	120 397	-173 500
3		187 353	0,7513	140 759	-32 741
4		229 018	0,683	156 420	123 679
5		270 683	0,6209	168 067	291 746

Рассчитаем РР для данного случая с учетом дисконтирования, используя формулу 4.3. Из приведенных в таблице 4.9 расчетов следует, что минимум непокрытого остатка инвестированной суммы дает 3-й год эксплуатационного периода (-69 454 руб.), следовательно,  $n_{цj} = 3$ . Тогда  $\frac{\Delta PR_{чj}}{PR_{чj+1}} = 32\,741 / 156\,420 = 0,21$ ; следовательно, РР равен 3,21 года.

Полученное значение говорит о том, что при таком же плане продаж, но с учетом изменения ценности денег, затраты на проект окупятся спустя 3 года и 2,1 месяца, что отличается на 4,7 месяца от расчетов без учета такой особенности. При этом, величина накопленного чистого эффекта, рассчитанная по формуле 5.4, равна:  $NPV = (62\,358 + 94\,568 + 120\,397 + 140\,759 + 156\,420 + 168\,067) - 450\,823 = 742\,569 - 450\,823 = 291\,746$  рубля.

Такие результаты говорят, что при следовании описанному плану продаж лицензий разработанного продукта и 5-летнем сроке реализации проекта, затраты на разработку не только полностью окупятся после 2,74 года на рынке,

но и к концу срока будет получен чистый доход, составляющий  $\frac{NPV}{I_0} = 291\,746 / 450\,823 = 64,7\%$  от суммы инвестиции в проект. Таким образом, расчеты сроков окупаемости подтверждают, что данный проект является экономически эффективным.

#### **4.4 Заключение по разделу**

В ходе выполнения магистерской диссертации была разработана информационная система поиска целевых сообществ в социальной сети для исследований в области общественно-политических наук, либо с маркетинговой целью.

В данном разделе было проведено планирование этапов научно-технического исследования, была рассчитана трудоемкость всех работ и занятость исполнителей при выполнении каждой из них. На основе этих данных были построены наглядные графики.

Также были выполнены расчеты полных денежных затрат на разработку проекта и определены сроки и условия его окупаемости. Как показало исследование, чем больше лицензий на программное обеспечение разработанной информационной системы будет продано в первый год на рынке, и чем дольше компании будут его использовать для решения задач мониторинга (продлевать действие лицензии), тем быстрее проект окупится и станет приносить прибыль. Приведенная оценка экономической эффективности является высокой и подтверждает целесообразности инвестиций в данный проект.

## **5 Социальная ответственность**

Данная магистерская диссертация направлена на создание алгоритмического и программного обеспечения (ПО), которое позволит осуществлять поиск целевых сообществ в социальной сети для социо-политических исследований и для целей маркетинга.

Областью применения разработки являются научно-исследовательские организации в области социо-политических исследований и предприятия маркетинговых услуг, осуществляющие поиск целевой аудитории в социальной сети для продвижения товаров и услуг.

Разработка осуществлялась за настольным персональным компьютером в офисе организации, специализирующейся на проектировании, производстве и внедрении высокоэффективных информационно-управляющих систем (ИУС) и заинтересованной в результатах по данной теме диссертации.

Так как основная работа при написании диссертации, разработка ПО, связана с компьютером, или персональной электронной вычислительной машиной (ПЭВМ), то в рамках текущего раздела целесообразным является рассмотрение следующих вопросов:

- выявление и изучение вредных и опасных производственных факторов при работе с ПЭВМ;
- определение способов снижения действия описанных факторов до безопасных пределов или по возможности до полного их исключения;
- безопасность окружающей среды;
- безопасность в чрезвычайных ситуациях (ЧС), которые могут возникнуть при эксплуатации ПЭВМ.

### **5.1 Правовые и организационные вопросы обеспечения безопасности**

Правовая сторона вопроса безопасности труда при работе над магистерской диссертацией регламентируется Трудовым кодексом Российской Федерации. Так количество рабочих часов в неделю не превышало 40 часов,

работа велась в будние дням, в течение рабочего дня предоставлялся часовой перерыв, который к рабочему времени не относится.

Деятельность при выполнении магистерской диссертации также связана с работой за компьютером (или ПЭВМ). Основным документом, регулирующим условия и организацию работы с ПЭВМ, является санитарно-эпидемиологические правила и нормативы СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы», которые включают ряд требований к ПЭВМ и организации рабочего места, а также к факторам, оказывающим на пользователя ПЭВМ опасное и вредное влияние [23]. Также используется ГОСТы регулирующие рабочее место пользователя ПЭВМ: ГОСТ 12.2.032-78 ССБТ «Рабочее место при выполнении работ сидя. Общие эргономические требования» и ГОСТ Р 50923-96. «Дисплей. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения» [24-25].

Для создания комфортной и безопасной среды для работы за компьютером необходимо учитывать эргономические требования к правильному расположению и компоновке рабочей зоны инженера. В офисном помещении ООО «ТомИУС-ПРОЕКТ», планировка рабочих мест учитывает требования к расстоянию между рабочими столами с видеомониторами и составляет 3 м. При этом расстояние от глаз до самого монитора составляет 650 мм, а конструкция ПЭВМ обеспечивает возможность поворота корпуса в горизонтальной и вертикальной плоскости с фиксацией в заданном положении для обеспечения фронтального наблюдения монитор. Т.е. приняты все меры по обеспечению оптимального размещения используемого оборудования на поверхности рабочего стола. Монитор также предусматривает регулирование яркости и контрастности, что уменьшает нагрузку на глаза и увеличивает комфортность работы на ПЭВМ.

Кроме того, форма рабочего стола должна быть удобна для поддержания рациональной позы пользователя, так, чтобы он мог менять положения своего

тела для предупреждения утомления. В таблицах 5.1-5.2 представлены фактические значения параметров рабочей зоны инженера в соответствии с требованиями СанПиН 2.2.2/2.4.1340-03, ГОСТ 12.2.032-78 ССБТ и ГОСТ Р 50923-96.

Таблица 5.1 – Анализ рабочего места

<b>Требование</b>	<b>Факт</b>
Окна должны быть ориентированы на север и северо-восток	<b>Все окна ориентированы на запад</b>
Оконные проемы должны быть оборудованы регулируемыми устройствами типа: жалюзи, занавесей, внешних козырьков	<b>Имеются жалюзи</b>
Площадь на одно рабочее место пользователей ПЭВМ на базе плоских дискретных экранов (жидкокристаллические, плазменные) - 4,5 м <sup>2</sup>	<b>На одно рабочее место выделена площадь 5 м<sup>2</sup></b>
Помещения, где размещаются рабочие места с ПЭВМ, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации	<b>Заземление имеется</b>
Не следует размещать рабочие места с ПЭВМ вблизи силовых кабелей и вводов, высоковольтных трансформаторов, технологического оборудования, создающего помехи в работе ПЭВМ	<b>Вблизи подобных объектов нет</b>

Анализ рабочего места на основе данных таблицы 5.1 показывает, что все требования к рабочему месту выполнены, кроме первого. Солнце вечером светит в окна, затрудняя работу, вследствие чего приходится использовать шторы и включать искусственное освещение.

Требования к столу и стулу, представленные в таблице 5.2, соответствуют требованиям СанПиН 2.2.2/2.4.1340-03 для людей ростом 161-175 см.

Таблица 5.2 – Анализ параметров рабочего стола и стула

<b>Требование</b>	<b>Допустимые значения</b>	<b>Фактическое значение</b>
Поверхность стола, мм	700	800
Пространство для ног, мм	640	700
Высота сиденья над полом, мм	420	450
Ширина сиденья, не менее, мм	340	405
Глубина сиденья, мм	380	380
Высота нижнего края спинки над сиденьем, мм	170	170
Высота верхнего края спинки над сиденьем, мм	360	360
Высота линии прогиба спинки, не менее, мм	210	215
Радиус изгиба переднего края сиденья, мм	20-50	45
Угол наклона сиденья	0-4°	3°
Угол наклона спинки	95-108°	90°
Радиус спинки в плане, не менее, мм	300	300

Фактические значения параметров рабочего стола и стула, как видно из таблицы 5.2, максимально приближены и соответствуют нормирующим требованиям. Однако стол, для людей с небольшим ростом, может оказаться высоким, в следствии чего посадка пользователя за ПЭВМ может оказаться неверной и приводить к повышенной утомляемости.

Кроме того, при работе с ПЭВМ было использовано рекомендуемое СанПиНом время перерывов 10-15 минут после каждых 45-60 минут работы с

использованием комплекса профилактических мероприятий: упражнения для глаз и физкультурные минуты. Также в помещении проводились ежедневные влажные уборки и систематическое проветривание каждый час.

## **5.2 Производственная безопасность**

В данном подпункте анализируются вредные и опасные факторы, влияющие на работу инженера на рабочем месте, оборудованном ПЭВМ.

Для идентификации потенциальных факторов использован ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» [26]. Все производственные факторы классифицируются по группам элементов: физические, химические, биологические и психофизические. Для данной работы целесообразно рассмотреть физические и психофизические вредные и опасные факторы производства, характерные как для рабочей зоны инженера-программиста, как разработчика рассматриваемой в данной работе системы. Так как психофизические факторы, связанные с нормированием рабочего дня и отдыха, а также с предоставлением комфортного рабочего места, были описаны в пункте 5.1 данного раздела, то далее будут рассматриваться физические факторы. Перечень данных факторов представлен в таблице 5.3.



Таблица 5.3 – Вредные и опасные производственные факторы при выполнении работ за ЭВМ

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Разраб отка	Экспл уатаци	
1. Отклонение показателей микроклимата	+	+	1. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.
2. Недостаточная освещенность рабочей зоны	+	+	
3. Превышение уровня шума	+	+	
4. Повышенный уровень электромагнитных излучений	+	+	2. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95*
5. Повышенное значение напряжение в электрической цепи, замыкание которой может произойти через тело человека	+	+	3. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности. 4. ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности. 5. ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.

В случае программной разработки, где объектом является рабочее место, включая персональный компьютер и помещение, среди таких вредных

воздействий можно указать: микроклимат помещения, неправильное освещение, шум, электромагнитное излучение, опасность поражения электрическим током и другие. Также немаловажно позаботиться о экологической безопасности и безопасности в чрезвычайных ситуациях, которые могут возникнуть при работе с ПЭВМ.

### 5.2.1 Микроклимат рабочего места

К параметрам микроклимата относятся: температура воздуха, температура поверхностей, относительная влажность воздуха, скорость движения воздуха.

Комплекс данных параметров воздействует на человека, определяя его самочувствие, а значит, и работоспособность. Оптимальные значения этих характеристик зависят от сезона (холодный, тёплый), а также от категории физической тяжести работы. Для инженера-программиста она является лёгкой (Ia), так как работа проводится сидя, без систематических физических нагрузок.

Учитывая вышеизложенное, можно привести набор оптимальных значений параметров микроклимата (табл. 5.4) в соответствии с СанПиН 2.2.4.548–96 [27] и фактические, которые были измерены на рабочем месте.

Таблица 5.4 – Параметры микроклимата в помещениях с использованием ПЭВМ

Период года	Категория работ	Температура воздуха, °С		Относительная влажность воздуха, %		Скорость движения воздуха, м/с	
		Оптим.	Факт.	Оптим.	Факт.	Оптим. м.	Факт.
Холодный	Ia	22-24	22-24	40-60	40-58	0,1	< 0,1
Теплый	Ia	23-25	21-24	40-60	50-56	0,1	< 0,1

Задача состоит в том, чтобы поддерживать эти параметры в оптимальном состоянии. Для этого предусмотрены следующие средства: центральное отопление, вентиляция (искусственная и естественная), искусственное

кондиционирование. Для отслеживания значений описанных параметров в офисе используется термометр, совмещенный с гигрометром. Показания данного прибора были в рамках определенных допустимых значений, следовательно, параметры микроклимата в офисе ООО «ТомИУС-ПРОЕКТ» соответствуют нормам СанПиН.

### **5.2.2 Освещенность рабочей зоны**

Так как работа инженера-программиста подразумевает зрительный тип работы, то организация правильного освещения имеет значительное место.

Пренебрежение данным фактором может привести к профессиональным болезням зрения. В рабочем помещении сочетаются естественное освещение (через окна) и искусственное освещение (использование ламп при недостатке естественного освещения).

Разряд зрительных работ программиста относится к категории III г (высокой точности), параметры искусственного освещения указаны в таблице 5.5 согласно актуализированной редакции СНиП 23-05-95 [28].

Таблица 5.5 – Нормативные значения освещенности

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подразряд зрительной работы	Контраст объекта с фоном	Характеристика фона	Искусственное освещение		
						Освещенность, лк		
						При системе комбинированного освещения		При системе общего освещения
						Всего	В том числе от общего	
Высокой точности	От 0,3 до 0,5	III	г	Средний и большой	Светлый и средний	400	200	200

Также согласно СанПиН 2.2.2/2.4.1340-03 уровень освещения на поверхности рабочего стола при работе с ПЭВМ должен быть в диапазоне от 300 до 500 лк.

Приведем расчет искусственного освещения для офиса организации, в котором осуществляется разработка программного обеспечения. Размеры офиса следующие: длина  $A = 4$  м, ширина  $B = 6$  м, высота  $H = 3$  м. В офисе используются светильники типа ОДР (светильник общего освещения диффузный с экранирующей решеткой) с люминесцентными лампы типа ЛБ (белый свет) мощностью 65 Вт и со световым потоком  $\Phi = 4600$  лм. Общее число ламп в офисе равно  $n = 8$ . Коэффициент пульсации ламп данного типа не превышает 5%, что соответствует нормам.

Освещенность помещения рассчитывается по формуле:

$$E_{\Phi} = \frac{n * \eta * \Phi}{S * k * z}, \quad (5.1)$$

где  $n$  – число светильников;

$\eta$  – коэффициент использования светового потока;

$\Phi$  – световой поток светильника, лм;

$S$  – площадь помещения,  $m^2$ ;

$k$  – коэффициент неравномерности освещения;

$z$  – это коэффициент неравномерности освещения.

Коэффициент запаса  $k$  учитывает запыленность светильников и их износ. Для помещений с вычислительной техникой  $k = 1,4$ . Поправочный коэффициент для люминесцентных ламп равен  $z = 1,1$ . Площадь помещения равна  $S = A * B = 5 * 6 = 24 m^2$ .

Коэффициент использования светового потока определяется при помощи таблицы на основе индекса помещения и коэффициенты отражения от стен, потолка и рабочей поверхности. Поэтому сначала найдем данные показатели.

Индекс помещения определяется по формуле:

$$i = \frac{S}{h * (A + B)}, \quad (5.2)$$

где  $S$  – площадь помещения,  $m^2$ ;

$A$  – длина комнаты, м;

$B$  – ширина комнаты, м;

$h$  – высота подвеса светильников, м.

При этом расчетная высота подвеса светильников над рабочей поверхностью ( $h$ ) в офисе компании определяется по формуле:

$$h = H - h_p - h_c,$$

где  $H$  – высота потолка в помещении, м;

$h_p$  – расстояние от пола до рабочей поверхности стола, м;

$h_c$  – расстояние от потолка до светильника, м.

Тогда расчетная высота подвеса светильников равна:

$$h = 3 - 0,8 - 0,01 = 2,19 \text{ м.}$$

Подставим полученное значение в формулу 5.2 для расчета индекса помещения.

$$i = \frac{S}{h*(A+B)} = \frac{4*6}{2.19*(4+6)} = 1,1$$

Исходя из того, что потолок в помещении чистый бетонный, стены бетонные оклеенные светлыми обоями с окнами и рабочая поверхность содержит ПЭВМ, то согласно [29], примем коэффициенты отражения от стен  $\rho_c = 30\%$ , потолка  $\rho_n = 50\%$  и от рабочей поверхности  $\rho_p = 10\%$ .

По таблице коэффициентов использования светового потока светильников с люминесцентными лампами [29] для соответствующих значений  $i$ ,  $\rho_c$ ,  $\rho_n$ , определяем коэффициент использования светового потока.

Получается для светильника ОДР при  $i = 1,1$   $\rho_c = 30\%$  и  $\rho_n = 50\%$  коэффициент использования светового потока равен  $43\%$ .

Учитывая все параметры, рассмотренные выше, найдем освещенность по формуле 5.1:

$$E_{\phi} = \frac{8 * 0,43 * 4600}{24 * 1,4 * 1,1} = 428 \text{ лк.}$$

В рассматриваемом помещении освещенность должна составлять не менее 300 лк согласно СанПиН 2.2.2/2.4.1340-03 и СНиП 23-05-95. В данном помещении освещенность равна 428 лк и находится в пределах нормы, следовательно, дополнительные источники света не нужны.

### 5.2.3 Шум

Источниками шума являются: работающее оборудование, вентиляторы компьютера, копировальная техника и кондиционеры.

Шум оказывает негативное воздействие на организм человека: снижает работоспособность, повышает утомляемость, воздействует на органы слуха и центральную нервную систему, снижает внимание.

Допустимый уровень шума ограничен ГОСТ 12.1.003-2014 ССБТ и СН 2.2.4/2.1.8.562–96. Уровень шума на рабочем месте программистов не должен превышать 50дБА, а в залах обработки информации на вычислительных машинах 65дБА [30-31].

Поэтому уровень шума в помещениях должен быть ограничен. Данные ограничения для требуемого помещения представлены в таблице 5.6.

Таблица 5.6 – Допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ

Уровни звукового давления в октавных полосах со среднегеометрическими частотами									Уровн и звук в дБА
31,5 Гц	63 Гц	125 Гц	250 Гц	500 Гц	1000 Гц	2000 Гц	4000 Гц	8000 Гц	
86 дБ	71 дБ	61 дБ	54 дБ	49 дБ	45 дБ	42 дБ	40 дБ	38 дБ	50

В рассматриваемом рабочем помещении уровень шума является допустимым и не превышает значений, установленных нормами и составляет не более 50 дБА. Кроме того, для уменьшения шума производятся: регулярное техническое обслуживание компьютеров и другой техники в помещении.

#### 5.2.4 Электромагнитное излучение

Персональные компьютеры являются источниками электромагнитных волн, то есть распространяющихся в пространстве возмущений электромагнитного поля (ЭМП). Все электрические приборы излучают такие волны, однако наибольший вклад вносит экран монитора. При определённых уровнях такие поля оказывают вредное влияние на человека: нарушение функционального состояния нервной и сердечно-сосудистой систем, это проявляется в повышенной утомляемости, понижении качества выполнения рабочих операций, изменении кровяного давления и пульса

Ввиду того, что используется жидкокристаллический монитор, то контроль мягкого рентгеновского излучения не осуществляется. Допустимые значения излучения показаны в таблице 5.7 с учётом СанПиН 2.2.2/2.4.1340-03 и ГОСТ 12.1.006-84 ССБТ [32].

Таблица 5.7 – Временные допустимые уровни (ВДУ) ЭМП, создаваемых ПЭВМ

Наименование параметров		ВДУ ЭМП	ВДУ ЭМП
Напряженность электрического поля	в диапазоне частот 5 Гц – 2 кГц	25 В/м	27 В/м
	в диапазоне частот 2 кГц – 400 кГц	2,5 В/м	2,5 В/м
Электростатический потенциал экрана видеомонитора		500 В	490 В

Нормы допустимых уровней напряженности электромагнитных полей зависят от времени пребывания человека в контролируемой зоне. Присутствие персонала на рабочем месте в течение 8 ч допускается при напряженности, не превышающей 5 кВ/м.

Основной способ снижения вредного воздействия – это увеличение расстояния от источника (не менее 50 см от пользователя). Защитой от воздействия электромагнитного поля токов промышленной частоты являются стационарные или переносные заземленные экранирующие устройства. На предприятии электромагнитное излучение не превышает 5 кВ/м, поэтому при работе за компьютером специальные экраны и другие средств индивидуальной защиты применены не были.

### 5.2.5 Опасность поражения током

Среди распространенных опасностей в рабочей зоне находится и поражение электрическим током. Опасность поражения определяется величиной тока проходящего через тело человека или напряжением прикосновения.



При получении человеком разряда электрического тока могут быть получены электротравмы, электрические удары и даже летальный исход. ГОСТ 12.1.038-82 ССБТ [33] определяет предельно допустимые значения напряжения прикосновения и тока на рабочем месте (см. табл. 5.8).

Таблица 5.8 – Допустимые значения напряжения прикосновения и тока

Род тока	Напряжения прикосновения, В	Ток, мА
	Не более	
Переменный, 50 Гц	2,0	0,3
Постоянный	8,0	1,0

Основным источником угрозы поражения электрическим током является персональный компьютер. Во избежание несчастных случаев сотрудники в обязательном порядке должны проходить соответствующий инструктаж.

Не следует работать на персональном компьютере при:

- повышенной влажности (относительная влажность воздуха более 75%);
- высокой температуре (более 35 °С);
- наличие токопроводящей пыли, токопроводящих полов и возможности одновременного соприкосновения к имеющим соединение с землёй металлическим элементам и металлическим корпусом электрооборудования.

Персональный компьютер питается от сети 220 В переменного тока с частотой 50 Гц. Это напряжение опасно для жизни, поэтому обязательны следующие меры предосторожности:

- перед началом работы нужно убедиться, что выключатели и розетка закреплены и не имеют оголённых токоведущих частей;
- при обнаружении неисправности оборудования и приборов необходимо, не делая никаких самостоятельных исправлений, сообщить человеку, ответственному за оборудование.

Чтобы избежать поражения электрическим током, необходимо защитить все токоведущие части от возможных прикосновений, а металлические корпуса должны быть заземлены.

Таким образом, все требования при работе с ПЭВМ были выполнены, так как все необходимые показатели норм находятся в допустимых пределах

#### **5.2.6 Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на работника**

Для снижения воздействия вредных и опасных факторов на работника на предприятии ООО «ТомИУС-ПРОЕКТ» проводится ряд профилактических действий на основе следующих рекомендаций.

##### ***Рекомендации по улучшению микроклимата***

К рекомендациям по оздоровлению воздушной среды на предприятии относятся правильная организация вентиляции (как естественным, так и механическим путем) и кондиционирования воздуха, отопление комнаты (в зимнее время года).

Объем помещений с ЭВМ не должен быть меньше 20 м<sup>3</sup>/человека [23].

##### ***Рекомендации по минимизации влияния шума***

Для того чтобы снизить шум следует:

- ослабить шум самих источников, используя звукоизоляцию;
- снизить эффект суммарного воздействия отраженных звуковых волн;
- использовать архитектурные и технологические решения, которые направлены на изоляцию источников шума.

##### ***Рекомендации по минимизации отрицательного влияния освещения рабочей зоны***

Для обеспечения требуемого уровня освещения в помещении используется лампы дневного освещения, равномерно распределенные по всему потолку офиса. Для освещения офисных помещений рекомендуется использовать лампы типа ЛБ (белый свет).

### ***Рекомендации по защите от электрического тока***

Нормативная база РФ устанавливает обязательные правила и меры безопасности во время работы с электрооборудованием.

Для защиты от поражения электрическим током все токоведущие части должны быть защищены от случайных прикосновений кожухами, корпус устройства должен быть заземлен. Питание устройства должно осуществляться от силового щита через автоматический предохранитель, который срабатывает при коротком замыкании нагрузки.

Для предотвращения поражения электрическим током в организации должны проводиться следующие мероприятия:

- компьютеры подключаются к сети с помощью трехполюсных вилок, причем центральный контакт вилки надежно заземляется;
- при эксплуатации электрооборудования рабочее место должно быть оборудовано так, что исключается возможность прикосновения служащих к токоведущим устройствам, шинам заземления, батареям отопления, водопроводным трубам;
- обслуживающий персонал должен пройти инструктаж по технике безопасности на рабочем месте;
- осуществляется профилактическая проверка отсутствия напряжения, отключение неисправного оборудования и наложение заземления.

### **5.3 Экологическая безопасность**

В настоящее время проблема экологической безопасности и охраны окружающей среды стоит под острым вопросом и является приоритетной. Это стало поводом для принятия жестких законов, ограничивающих обычную утилизацию компьютерной техники. В большей мере это обуславливается тем, что в производстве такой техники используется множество различных материалов, которые способны нанести непоправимый вред окружающей среде и, соответственно, здоровью человека. Утилизация компьютерного оборудования является достаточно сложной. Непосредственная переработка

большей части компонентов включает в себя их сортировку, последующую гомогенизацию и отправку для повторного использования, т.е. с предварительным помолом или переплавкой. Люминесцентные лампы представляют собой «чрезвычайно опасные» виды отходов [9]. Содержание ртути в любых люминесцентных лампах составляет от трех до пяти миллиграмм ртути. С учетом этого необходимо обеспечивать определенные условия хранения, их эксплуатации и утилизации. Согласно санитарным нормам хранить ртутесодержащие отходы необходимо в специальных герметичных контейнерах, доступ посторонним лицам к таким контейнерам должен быть запрещен.

Транспортировка ламп на полигоны складирования должна выполняться организациями, которые специализируются на утилизации опасных отходов. Категорически запрещено размещение таких отходов, как люминесцентные лампы на полигонах твердых бытовых отходов.

Так как при разработке данной магистерской диссертации использовался персональный компьютер, необходимо описать правильную утилизацию компьютерного лома после его выхода из строя. В соответствии с постановлением правительства юридическим лицам запрещено самостоятельно утилизировать компьютерную технику. Для этого необходимо найти специальную компанию, которая занимается утилизацией в частном порядке.

В нормативном документе СанПиН 2.2.2/2.4.1340-03 даются следующие общие рекомендации по снижению опасности для окружающей среды, исходящей от компьютерной техники:

- применять оборудование, соответствующее санитарным нормам и стандартам экологической безопасности;
- применять расходные материалы с высоким коэффициентом использования и возможностью их полной или частичной регенерации;
- отходы в виде компьютерного лома утилизировать;
- использовать экономичные режимы работы оборудования.

## 5.4 Безопасность в чрезвычайных ситуациях

Самым распространенным чрезвычайным обстоятельством в офисе является пожар. Такое рабочее место относится к категории «В» (пожароопасные), так как в данном помещении присутствует пыль, вещества и материалы, способные при взаимодействии с воздухом гореть.

Возникновение пожара может произойти по нескольким факторам:

- возникновением короткого замыкания в электропроводке вследствие неисправности самой проводки или электросоединений и электрораспределительных щитов;
- возгоранием устройств вычислительной аппаратуры вследствие нарушения изоляции или неисправности самой аппаратуры;
- возгоранием мебели или пола по причине нарушения правил пожарной безопасности, а также неправильного использования дополнительных бытовых электроприборов и электроустановок;
- возгоранием устройств искусственного освещения.

Методы борьбы с пожарами предусматривают:

- Инструктажи, наличие планов эвакуаций, правильный монтаж и эксплуатация оборудования, правильное содержание зданий и территорий, обучение правилам техники безопасности, издание специальных инструкций и плакатов
- Соблюдение противопожарных правил, исключение образования горючей среды, применение трудно сгораемых материалов
- Предусмотренные средства сигнализации, огнетушители, автоматические стационарные системы тушения пожаров, своевременная эвакуация.

Для предотвращения пожара помещение с ПЭВМ должно быть оборудовано первичными средствами пожаротушения: углекислотным огнетушителем типа ОУ-2 или ОУ-5 [34].

Пожар может нанести не только вред здоровью, но и материальный ущерб. Применимо к выполняемой работе в случае пожара могут быть уничтожены бумажные документы и\или электронные носители информации.

Для защиты информации рекомендуется использовать облачные хранилища данных для данных и документов. Для исходных кодов программ рекомендуется использовать системы контроля версий.

## **5.5 Заключение по разделу**

Проанализировав условия труда на рабочем месте можно сделать вывод о том, что помещение, в котором ведётся разработка программного обеспечения, удовлетворяет необходимым нормам и в случае соблюдения техники безопасности и правил пользования компьютером работа в данном помещении не приведет к ухудшению здоровья.

Помещение и рабочее место удовлетворяет всем нормативным требованиям. Во избежание негативного влияния на здоровье во время работы с ПЭВМ необходимо делать перерывы и проводить специализированные комплексы физических упражнений.

Действие вредных и опасных факторов сведено к минимуму соответствующими мероприятиями.

Микроклимат, освещение и электро- и пожаробезопасность соответствуют требованиям, предъявленным в соответствующих нормативных документах.

Экологическая безопасность на предприятии удовлетворяет действующим нормативным документам.

## **Список используемых источников:**

1. Коршунов А., Белобородов И., Бузун Н., Аванесов В., Пастухов Р., Чихрадзе К., Козлов И., Гомзин А., Андрианов И., Сысоев А., Ипатов С., Филоненко И., Чуприна К., Турдаков Д., Кузнецов С. Анализ социальных сетей: методы и приложения. Труды Института системного программирования РАН. 2014;26(1):439-456.
2. И. Г. Чередов. Анализ целевой аудитории в интернете: Алгоритмы и инструменты. Стратегические коммуникации в бизнесе и политике. Материалы международной научной конференции 23-24 ноября 2016г. №2. г. Санкт-Петербург. 100-108.
3. Ю.А. Брильков. Автоматизация поиска и аналитики целевой аудитории в социальных сетях. Новые направления маркетинговой политики хозяйствующих субъектов, г. Нижний Новгород, 05 июня 2015 г., 69-73.
4. Гойко В.Л., Киселёв П.Б., Мацута В.В., Суханова Е.А., Сепаненко А.А., Фещенко А.В. Методы и инструменты выявления перспективных абитуриентов в социальных сетях. Открытое и дистанционное образование. 2017. № 4(68). DOI: 10.17223/16095944/68/7.
5. О. О. Слабченко, В. Н. Сидоренко, Р. А. Пономарчук. Методы и алгоритмы выявления сообществ потенциальных абитуриентов и их лидеров в социальных сетях. Вестник КрНУ имени Михаила Остроградского. Выпуск 1/2013 (78).
6. Dmitry Donchenko, Nadezhda Ovchar, Natalia Sadovnikova, Danila Parygin , Olga Shabalina, Danish Ather. Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension. 6th International Young Scientists Conference in HPC and Simulation, YSC 2017, 1-3 November 2017, Kotka, Finland.
7. Vasiliy A. Perepelitsyn, Alla G. Kravets. The social networks' nodes grouping algorithm for the analysis of implicit communities. 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA).

8. Michele Coscia, Fosca Giannotti, Dino Pedreschi. A Classification for Community Discovery Methods in Complex Networks. *Statistical Analysis and Data Mining journal*, Special Issue: Networks. Volume 4, Issue 5, pages 512-546, October 2011.
9. Matthew Curran Benigni. Detection and Analysis of Online Extremist Communities. Institute for Software Research, School of Computer Science, Carnegie Mellon University Pittsburgh.
10. Santo Fortunato. Community detection in graphs. *Physics Reports* 486 (2010) 75–174.
11. Rumi Ghosh, Kristina Lerman. Community Detection Using a Measure of Global Influence. USC Information Sciences Institute, Marina del Rey, California 90292.
12. M. I. Kolomeychenko, A. A. Chepovski, A. M. Chepovski. An algorithm for detecting communities in social networks. *Journal of Mathematical Sciences*, Vol. 211, No. 3, December, 2015.
13. Liu, W., Pellegrini, M. & Wang, X.-F. Detecting Communities Based on Network Topology. *Sci. Rep.* 4, 5739; DOI:10.1038/srep05739 (2014).
14. Chen, Y., Zhao, P., Li, P. et al. Finding Communities by Their Centers. *Sci Rep* 6, 24017 (2016). <https://doi.org/10.1038/srep24017>.
15. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu. *Social Media Mining*. Cambridge University Press. Draft Version: April 20, 2014.
16. Lei Tang, Huan Liu. Community Detection and Mining in Social Media. *Community Detection and Mining in Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery* 2010, 137 pages.
17. Savelev A., Karpova A., Chaykovskiy D., Vilnin A., Kaida A., Kuznetsov S., Igumnov L., Maksimova N. The High-Level Overview of Social Media Content Search Engine. *Proceedings of the 14th International Forum on Strategic Technology (IFOST 2019)*; Tomsk Polytechnic University. – Tomsk: TPU Publishing House, 2019. – pp. 306-309.



18. Анализ способов сбора социальных данных из сети интернет. Суханов А. А., Маратканов А. С. МГТУ им. Н.Э. Баумана. [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/analiz-sposobov-sbora-sotsialnyh-dannyh-iz-seti-internet> (дата обращения: 26.06.2019).
19. Django. [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/Django> (дата обращения: 01.06.2020).
20. HTML. [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/HTML> (дата обращения: 01.06.2020).
21. CSS. [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/CSS> (дата обращения: 01.06.2020).
22. BootStrap. [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/Bootstrap\\_\(%D1%84%D1%80%D0%B5%D0%B9%D0%BC%D0%B2%D0%BE%D1%80%D0%BA\)](https://ru.wikipedia.org/wiki/Bootstrap_(%D1%84%D1%80%D0%B5%D0%B9%D0%BC%D0%B2%D0%BE%D1%80%D0%BA)) (дата обращения: 01.06.2020).
23. СанПиН 2.2.2/2.4.1340–03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.
24. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
25. ГОСТ Р 50923-96. Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения.
26. ГОСТ 12.0.003-2015. Опасные и вредные производственные факторы. Классификация.
27. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений.
28. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95\*.
29. Безопасность жизнедеятельности: практикум / Ю.В. Бородин, М.В. Василевский, А.Г. Дашковский, О.Б. Назаренко, Ю.Ф. Свиридов, Н.А. Чулков, Ю.М. Федорчук. — Томск: Изд-во Томского политехнического университета, 2009. — 101 с.

- 30. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности.
- 31. СН 2.2.4/2.1.8.562–96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории застройки.
- 32. ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля (с Изменением N 1).
- 33. ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов (с Изменением N 1).
- 34. ГОСТ Р 22.3.03-94. Безопасность в ЧС. Защита населения. Основные положения.

## Приложение А (справочное)

### Chapter 1

#### Analytical review of scientific and technical documentation

Студент:

Группа	ФИО	Подпись	Дата
8ИМ82	Кузнецов Сергей Анатольевич		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

Консультант-лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Маркова Наталия Александровна			

## **Introduction**

In 2018 regular studies were conducted on active audiences of Russian social networks. Studies included the data gathered around such social networks as VKontakte (VK), Odnoklassniki (OK), Facebook, Instagram, YouTube, Twitter, My world and LiveJournal.

The attention of studies was focused on active (posting) audience as long as social networks are studied as a way of public communication and as an influencer on the public opinion.

The data used during studies include such categories as audience, age, sex and regional distribution of the posting authors of Russian social network. The data also include the most popular authors and groups of the most popular networks.

Social media in Russia can be presented by the quantity of written messages per one month which is 1,8 billion and, by the quantity of authors which is 46 million.

The common trend for all social networks is adulting. The most visible example is Facebook whose audience growth was provided by users over 45 years. The expulsion from the row is Twitter explored by the young generation and OK presenting the attraction for younger people content.

The problem of algorithmic news feed, advertising noise and informational chocking are still real and that is the reason why users pay more attention to the other platforms for increasing of communication qualities alongside with decrease of consumable amount of information. Message boards and topical platforms, social network communities and such platforms which allow us to configurate our own personal media zone play such a role firstly. For example, Twitter uses hashtags, Telegram uses public channels and chats and etc. Similar to the record growth on the world stage of the Reddit project, the algorithmic personal news channel Yandex.Zen is growing in Russia, whose users show high activity in commenting on the site itself and broad citation outside it. Over the year, Zen outcompeted in citing social media leaders such as RIA Novosti and TASS.

The key goal of social networks is to improve the quality of media consumption and to retain the user, to offer new formats and ways of earning money to both authors

and professional content providers who themselves could not cope with the distribution of content outside of TV. The healthy competition between the platforms for authors plays into the hands of authors, media providers and users alike.

Consequently, we can conclude that social networks and services are gaining more weight as a source of data for sociological, political and other types of studies of the audience.

The aim of the work is to develop algorithms and corresponding software to increase the efficiency of search and subsequent analysis of target communities in social networks.

## **1. Analytical review of scientific and technical documentation**

### **1.1 Methods and solutions for finding targeted communities in social networks**

The studies of Russian authors in the field of analysis of social networks are devoted to the description of the main components of the developed technology stack for analyzing user data from a social network [1]. Researches point at the study of popular practical tools and services for analyzing data on the behavior of the target audience on the Internet [2]. The problem of automating the collection and analysis of the target audiences in social networks using as the example of the most functional software solution for SMM at the moment is the Cerebro Target program which also finds its place in intellectual discourse [3]. The authors of [4] presented the experience of Tomsk State University in determining the target model of the applicant on the basis of data from the VK social network and developed tools for uploading and data mining and identifying gifted high school students for entering the university [4]. Data Mining as a concept for identifying communities of potential applicants in social networks and their leaders was proposed. It is a software application that collects information from social networks as part of constructing an experimental model for searching communities and their leaders has been implemented [5].

The authors [2, 3] use existing parsing services for study, the authors [1, 4, 5] developed their tools using the software interface of the VK API application.

As part of the study [4], a hypothesis of the relationship between the interests of students in content on social networks and the choice of a vocational training program was tested. The frequency analysis of subscriptions to the thematic VK communities allowed us to model the profile of the student's interests in the social network according to a simple classification: human sciences, exact sciences and natural sciences interests. The forecast accuracy was 0.82 for the human sciences interests, 0.76 for the exact sciences interests, and 0.69 for the natural sciences interests. The hypothesis that the personal page texts can express the scientific and professional interests of the applicant was also confirmed.

Studies [1, 4, 5] use information about social connections of users, that is a list of friends to search for targeted communities.

Methods and algorithms for searching target communities proposed in [5], present an interesting view on the structure of targeted search. The authors propose to consider a community of applicants in a different from the traditional paradigm, as a complex interconnected multifaceted structure that develops as a whole, with its explicit and implicit connections, leaders and dynamics. Such a view actualizes the need to develop the concept of an information system that allows you to search, collect and enrich information about potential applicants from a social network, visualize the connections between them and identify their communities and leaders. The proposed system is a direct marketing solution that includes the steps of collecting, pre-processing, analyzing and interpreting the results obtained with the need to further develop a specific information impact strategy.

The starting points for collecting information and building a model are indicators of the geographical position of the university and the task of the search radius. The data collected from the social network have the following characteristic features: high dimensionality, large volume and the presence of gaps. Therefore, the problem of their justification, selection and reduction before starting to build models arises. The problem is solved by applying a sequential reduction of the source data, i.e. the allocation of target groups of users of the social network with the subsequent clustering of the graph of links between the selected users (search for explicit and implicit

communities). As a result, the groups of the most probable applicants are distinguished, within which there is a search for leaders (the most influential agents).

In order to identify the most active participants in the target community from the point of view of their information interaction with other users, a graph of relationships for certain target groups is constructed.

Leaders identification of the applicant's community is provided by calculating the value of Betweenness Centrality for each vertex. When searching for community leaders, this algorithm requires setting a threshold value at which vertices with such and a large value are identified as leaders.

Thus, the most part of the studies presented in the Russian-language field is aimed at using already developed parsing services as a data source, and only a small fraction of the research concerns directly the processes of data collection, pre-processing, analysis and interpretation of data. At the same time, studies that include a complete list of tasks relate to highly specialized tasks, in particular, the selection of applicants for admission to the university.

In English sources the works in the field of analysis of social networks are devoted to the development of an approach for measuring social tension in the certain regions of Russia, based on the analysis of user messages in the VK social network [6]. Researches point at discussion of methods and algorithms for identifying implicit communities, that is virtual communities with similar interests [7]. Studies are focused on creating a guide on community discovery [8], methods for discovering large dynamic online activists and gaining a new understanding of major online activists [9]. The next works are dedicated to community searches using social graph analysis methods [10], to the development of community search algorithms using measurement results [11] and, to the development of simple search algorithms for new communities [12]. Finally, research activities are concentrated on the proposal of a simple and new structures for detecting communities based on network topology [13] and, on development of a simple approach that simultaneously reveals communities and their centers [14].

Thus, the major part of the research presented in the English-language field is aimed at using information from social connections of users to search for targeted communities.

As a result of the analytical review, we can conclude that the problem of increasing the accuracy of the search for target communities should be considered relatively to specific tasks. There is a need to develop methods for increasing the relevance of the search for targeted communities in a social network using the characteristic features of communities.

## **1.2 Features and limitations of the process of communities' analysis in social networks**

One of the most commonly used methods is obtaining data through application programming interfaces (APIs) from social media sites. Despite the convenience of using the API, there is one limitation: a social network cannot give away all the data that is visible to users in the interface. Firstly, this is connected with the fact that social networks try to maintain the privacy of their users, and secondly, to the fact that some functions load the server side of the application too much.

Unaware of the distribution of the population, we cannot know that our data samples are reliable representatives of complete data. Therefore, we cannot guarantee that our results obtained on social networks are reliable.

In the classic data mining literature, successful data acquisition involves extensive data preprocessing and noise removal. By nature, social media data can be most of the noise. For this data, the following can be noted: blind removal of noise can worsen the problem, because removal can also eliminate valuable information, and the determination of noise becomes complex and relative due to depending on the task.

The standard procedure for evaluating patterns in the data obtained should be of practical use. For example, a data set can be divided into training and test sets. Only training data is used in training, and test data is example for testing. However, the practical benefits are often not available in the data obtained from social networks.

The amount of data received from social networks is huge. Nevertheless, when we approach to the specific user for whom, for example, we would like to draw the



certain conclusions, we often have not enough of data. We must use the characteristics of social networks and use their multidimensional, multi-source and multi-node data to aggregate information with sufficient statistics to effectively obtain data.

Social networks allow new laboratories to study human relationships. These large networks, combined with unique features, create new challenges in data acquisition.

Some examples are given below.

**Scalability.** The network presented in social media can be huge, often in a scale of millions of actors and hundreds of millions of connections, while traditional social network analysis normally deals with hundreds of subjects or fewer. Existing network analysis techniques might fail when applied directly to networks of this astronomical size.

**Heterogeneity.** In reality, multiple relationships can exist between individuals. Two persons can be friends and colleagues at the same time. Thus, a variety of interactions exist between the same set of actors in a network. Multiple types of entities can also be involved in one network. For many social bookmarking and media sharing sites, users, tags and content are intertwined with each other, leading to heterogeneous entities in one network. Analysis of these heterogeneous networks involving heterogeneous entities or interactions requires new theories and tools.

**Evolution.** Social media emphasizes timeliness. For example, in content sharing sites and blogosphere, people quickly lose their interest in most shared contents and blog posts. This differs from classical web mining. New users join in, new connections establish between existing members, and senior users become dormant or simply leave. How can we capture the dynamics of individuals in networks? Can we find the die-hard members that are the backbone of communities? Can they determine the rise and fall of their communities?

**Collective Intelligence.** In social media, people tend to share their connections. The wisdom of crowds, in forms of tags, comments, reviews, and ratings, is often accessible. The meta information, in conjunction with user interactions, might be useful

for many applications. It remains a challenge to effectively employ social connectivity information and collective intelligence to build social computing applications.

Evaluation. A research barrier concerning mining social media is evaluation. In traditional data mining, we are used to the training-testing model of evaluation. It differs in social media. Since many social media sites are required to protect user privacy information, limited benchmark data is available. Another frequently encountered problem is the lack of ground truth for many social computing tasks, which further hinders some comparative study of different works. Without ground truth, how can we conduct fair comparison and evaluation.

The main tasks of the analysis of social networks include:

### 1) Network Modeling

Network modeling offers a deep understanding of network dynamics that is independent of network domains. The network model can be used to simulate various network properties, for example, the stability of the attacked network, the dissemination of information within a given network structure, etc.

When networks scale to millions or more nodes, it becomes a problem to compute some network statistics, such as diameter and average clustering coefficient.

Recently, the methods of using the capabilities of distributed computing have attracted increasing attention.

### 2) Centrality analysis and impact modeling

The analysis of centrality consists in determining the most “important” nodes in the network.

A traditional analysis of social networks depends on the link structure for identifying nodes with high centrality. Usually the centrality criterion is used.

On social networks, additional information is available, such as comments and tags. This information makes it possible to combine various sources of information for the study of centrality.

A related task is modeling of influence, which is aimed at understanding the process of influence. A subset of nodes that maximizes impact on the population.

### 3) Community Discovery

Communities are also called groups, clusters, connected subgroups, or modules in different contexts. This is one of the fundamental tasks in the analysis of social networks. In fact, the “founders” of sociology argued that the causes of social phenomena should have been found by studying groups, not individuals. To find a community in a social network means to define a set of nodes, since they interact with each other more often than with nodes outside the group.

Community discovery can facilitate other social networking tasks. It is used in many essential applications. For example, a group of customers with similar interests on social networks provides effective recommendations that open up a wide range of relevant items to customers to improve their shopping experience. Communities can also be used to compress a huge network into a smaller network. In other words, the solution to the problem is carried out at the group level, not in the network.

#### 4) Classification and recommendation

Classification and recommendations tasks are common in applications for social networks. A successful social networking site is often in demand in a large number of users. It is very important for the site to provide individual recommendations (e.g. friends, communities, tags) in order to stimulate more user interactions.

Classification may help recommendations. For example, one common feature in many social networks is the recommendation of friends, which offers a list of friends that the user may know.

Since it is difficult for a user to determine who is on a social network, this is an effective way to expand their connections.

#### 5) Confidentiality, spam and security

Social media lead to the socialization of people.

Many social networking sites (e.g. Facebook, Google Buzz) are often in heated debate about user privacy. This is also a difficult problem to be solved.

Spam is another issue of major concern on social networks.

The media often includes a large amount of private information. Some spammers use fake identifiers to obtain personal information from users on social networking sites.

Thus, mining in social networks is a young and dynamic direction with great opportunities. Social networks surprise us with their new forms and diversity. An interesting trend is that social networks increasingly get into the physical world with the latest mobile technologies and smartphones. In other words, as social media is more and more attracted to people's daily lives, this gap between the virtual and physical worlds is disappearing, which is a harbinger of the convergence of social network data analysis algorithms and real data analysis algorithms.

### **1.3 Specification of requirements for the developed algorithms and software**

Basic requirements for the developed software:

- There should be an automated search function for targeted communities by keywords and related communities.
- There should be a function of finding connections and mutual influence of communities.
- There should be a function for evaluating the calendar activity of the target community.
- There should be a function for visualizing community links and statistical information.

Description of the main interfaces and use cases

#### *Authorization Requirements*

On the authorization page, the user should have access to:

- Enter login values and the corresponding password.
- Go to the new user registration page.
- Go to the password recovery page.

#### *Registration Requirements*

On the registration page, the user must fill in the required fields: account name, password (and password confirmation), contact email.

#### *Default Page Requirements*

Must contain a general description of the features of the application

#### *Search*

*Search Requirements. KW&KD (keywords and dates)*

The following actions should be available on the page for an authorized user:

- Manual entry of keyword and phrases to search for target communities.
- Selecting a previously created KW&KD as a source for community searches.
- Import an excel file with a set of KW&KD as source data.
- Setting values for additional filters for search results (number of community users, activity dates).
- Editing search results.
- Save your search results as a new community list.

#### *Related Community Search Requirements*

The following actions should be available on the page for an authorized user:

- Manual entry of community identifiers to search for related communities.
- Selection of a previously created list of communities as input to search for related ones.
- Import of an Excel file with a list of communities as source data.
- Setting values for additional filters for search results (number of community users, activity dates).
- Editing search results.
- Save your search results as a new community list.

#### *Knowledge base*

##### *Knowledge Base Requirements. KW&KD*

The following actions should be available on the page for an authorized user:

- Create a new KW&KD set with manual data entry.
- Create a new KW&KD set based on an existing one.
- Create a new set of KW&KD based on the imported excel file.
- Editing the selected KW&KD set.
- Export of the selected KW&KD set to an excel file.

##### *Knowledge Base Requirements. Communities*

The following actions should be available on the page for an authorized user:

- Creating of a new community list with manual data entry.

- Creating of a new community list based on the existing ones.
- Creating of a new list set based on the imported Excel file.
- Editing the selected community list.
- Export of the selected community list set to an Excel file.

### *Analysis*

#### *Analysis Requirements. CAC*

On the page, an authorized user should have access to:

a) input of input data for analysis of:

- community list;
- manually formed;
- imported from an Excel file;
- previously created;
- previously created KW&KD kit;

b) creating a task to perform.

#### *Analysis Requirements. Geolocation*

On the page, an authorized user should have access to:

a) input of the initial data for the analysis of the list of communities:

- manually formed;
- imported from an Excel file;
- previously created.

b) creating a task to perform.

#### *Analysis Requirements. Social Graph*

On the page, an authorized user should have access to:

a) input of input data for analysis:

- community list;
- manually formed;
- imported from an Excel file;
- previously created;

- the choice of the graph construction method (by community links or by common users).
- b) creating a task to perform.

#### **1.4 Review conclusions**

An analytical review of the subject area, including a review of existing software for searching targeted communities on social networks, allows us to conclude that existing web services do not allow finding communities related to a specific subject area, and there are no tools for analyzing communities and users social network.

Based on the review and analysis of the subject area, we can formulate the goal of this work as follows: to develop algorithms and software to improve the efficiency of search and subsequent analysis of target communities in social networks.

## Приложение Б



Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

### АКТ ВНЕДРЕНИЯ результатов выпускной квалификационной работы студента группы 8ИМ82 Кузнецова Сергея Анатольевича

Следующие результаты магистерской диссертации Кузнецова С.А. были использованы при выполнении работ по проектам РФФИ 19-011-31535 «Новые коммуникативные технологии в продвижении ультраправой идеологической платформы среди молодежи: методика автоматизированного выявления угроз методами web mining» и ВИУ-УОД-145/2019 «Создание прототипа системы интеллектуального контент-анализа Интернет-ресурсов для диагностики и профилактики радикализации студентов высших учебных заведений»:

- алгоритмическое и программное обеспечение параметризованного поиска целевых сообществ в социальной сети;
- алгоритмическое и программное обеспечение автоматизированного формирования профиля целевого сообщества в социальной сети;
- алгоритмическое и программное обеспечение визуализации геолокационных данных участников сообществ социальной сети.

Полученные результаты позволяют автоматизировать решение задач сбора и предварительной обработки данных социальных сетей при проведении исследований в области общественно-политических наук.

Руководитель проектов, д.соц.н.,  
доцент ОСГН ШБИП

Директор ШБИП



А.Ю. Карпова

Д.В. Чайковский

А.О. Савельев  
ин. 4413