

Инженерная школа информационных технологий и робототехники
 Направление подготовки: 09.03.04 «Программная инженерия»
 Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

Тема работы Алгоритмическое и программное обеспечение поиска и анализа данных социальных сетей
--

УДК 004.4.021:004.6:316.472.4

Студент

Группа	ФИО	Подпись	Дата
8K51	Лебедева Яна Евгеньевна		

Руководитель ВКР

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

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСТН ШБИП ТПУ	Подопригора Игнат Валерьевич	К.Э.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП ТПУ	Винокурова Галина Федоровна	К.Т.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР ТПУ	Чердынцев Евгений Сергеевич	К.Т.Н.		

**Планируемые результаты обучения по направлению 09.03.04
«Программная инженерия»**

Код результата	Результат обучения (выпускник должен быть готов)
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т. п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования.
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки: 09.03.04 «Программная инженерия»
Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

бакалаврской работ

Студенту:

Группа	ФИО
8К51	Лебедевой Яне Евгеньевне

Тема работы:

Алгоритмическое и программное обеспечение поиска и анализа данных социальных сетей	
Утверждена приказом директора (дата, номер)	№1513/с от 26.02.2019

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

<p>Исходные данные к работе</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>	<p>Объектом исследования в данной работе является социальная сеть ВКонтакте;</p> <p>Режим работы: непрерывный;</p> <p>Особые требования к продукту: собранные данные не должны нарушать законы РФ;</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</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">1. Исследование предметной области2. Аналитический обзор существующих средств для обработки и визуализации данных3. Проектирование информационной системы по обработке и визуализации данных из социальных сетей4. Разработка информационной системы по обработке и визуализации данных из социальных сетей5. Финансовый менеджмент6. Социальная ответственность

Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	<ol style="list-style-type: none"> 1. Модель работы программы (Диаграмма в нотации IDEF0) 2. Диаграмма EPC 3. Диаграмма BPMN процесса обработки и визуализации данных 4. Диаграмма вариантов использования 5. Рисунки, демонстрирующие результаты 6. Диаграмма Ганта
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Подопригора Игнат Валерьевич
Социальная ответственность	Винокурова Галина Федоровна
Названия разделов, которые должны быть написаны на русском и иностранном языках:	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель:

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
8К51	Лебедева Яна Евгеньевна		

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники
Направление подготовки: 09.03.04 «Программная инженерия»
Уровень образования бакалавриат
Отделение информационных технологий
Период выполнения осенний/весенний семестр 2018/2019 учебного года

Форма представления работы:

бакалаврская работа

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
	Основная часть	60
	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	20
	Социальная ответственность	20

Составил преподаватель:

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

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР ТПУ	Чердынцев Евгений Сергеевич	к.т.н.		

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

Студенту:

Группа	ФИО
8К51	Лебедевой Яне Евгеньевне

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.04 Программная инженерия

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Амортизационные затраты на спецоборудование – 4769 рублей; Затраты на основную и дополнительную з/п – 125913 + 18886,95 рублей;
2. <i>Нормы и нормативы расходования ресурсов</i>	Затраты на отчисление во внебюджетные фонды – 40543 рубля; Накладные расходы – 30417,9 рубля.
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	Описание потенциальных потребителей; Технология QuaD; SWOT-анализ.
2. <i>Планирование и формирование бюджета научных исследований</i>	Структура работ в рамках научного исследования Определение трудоемкости выполнения работ и разработка графика проведения научного исследования Бюджет проекта
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	Определение интегрального финансового показателя разработки Определение интегрального показателя ресурсоэффективности разработки Определение интегрального показателя эффективности

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Оценка конкурентоспособности технических решений</i>
2. <i>Матрица SWOT</i>
3. <i>График проведения и бюджет НИ</i>
4. <i>Оценка ресурсной, финансовой и экономической эффективности НИ</i>

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСТН ШБИП ТПУ	Подопригора Игнат Валерьевич	Кандидат экономических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8К51	Лебедева Яна Евгеньевна		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8К51	Лебедевой Яне Евгеньевне

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.04 Программная инженерия

Тема ВКР:

Алгоритмическое и программное обеспечение поиска и анализа данных социальных сетей	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования – приложение для сбора и обработки данных из социальных сетей. Рабочее место – рабочий стол с персональным компьютером в учебной аудитории.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	- Рабочее место соответствует ГОСТу 12.2.032 – 78 - Организация рабочих мест с ЭВМ регулируется СанПиНом 2.2.2/2.4.1340 – 03 - Правовые нормы при сборе данных контролируются правилами пользования сайтом ВКонтакте
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<ul style="list-style-type: none"> • Отклонение показателей микроклимата • Отсутствие или недостаток естественного света • Недостаточная освещенность рабочей зоны • Повышенный уровень электромагнитных излучений
3. Экологическая безопасность:	Негативное влияние на окружающую среду заключается только в утилизации оргтехники, люминесцентных ламп, батареек и макулатуры.
4. Безопасность в чрезвычайных ситуациях:	Возможные чрезвычайные ситуации: пожар
Дата выдачи задания для раздела по линейному графику	

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП ТПУ	Винокурова Галина Федоровна	Кандидат технических наук, доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8К51	Лебедева Яна Евгеньевна		

Реферат

Выпускная квалификационная работа: 65 с, 14 рисунков, 20 таблиц, 7 источников, 1 приложение.

Ключевые слова: социальные сети, анализ, визуализация.

Объектом исследования является социальная сеть ВКонтакте.

Предмет исследования: анализ и визуализация данных.

Цель работы – разработка алгоритмического и программного обеспечение анализа и визуализации данных социальных сетей.

В процессе исследования проводился аналитический обзор существующих средств анализа социальных сетей и визуализации данных, были определены вопросы исследования и проведены анализ и визуализация результатов.

В результате исследования было выявлено, что большинство студентов Томского Политехнического Университета предпочитает не указывать личную информацию в социальных сетях, поэтому, для получения более точных результатов, необходимо анализировать неструктурированную информацию.

Экономическая эффективность работы в высокой научной значимости и актуальности поставленной задачи.

В дальнейшем планируется проводить более детальный и глубокий анализ не только студентов Томского Политехнического Университета, но и его сотрудников, а также потенциальных абитуриентов.

Список терминов и сокращений

В данной работе используются следующие термины:

Социальная сеть – ресурс, предназначенный для обеспечения взаимодействия между людьми либо организациями в Интернете.

Контент – содержимое, информационное наполнение.

Методология IDEF0 – метод описания и формализации бизнес-процессов.

Проектирование информационной системы – это поиск способа, удовлетворяющего требованиям функциональности системы имеющимися средствами с учетом заданных ограничений.

Событийная цепочка процессов (EPC) – тип блок-схемы, используемой для бизнес-моделирования.

BPMN (The Business Process Modeling/Management Notation) – система условных обозначений для построения схемы протекания бизнес-процессов.

Нормативные ссылки

В данной работе используются ссылки на следующие стандарты:

ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация.

ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.

ГОСТ 12.1.038–82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.

ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.

ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования.

ГОСТ 19.402-78. ЕСПД. Описание программы.

Содержание

Введение	13
Глава 1. Исследование предметной области	16
1.1. Постановка задачи	16
1.2. Описание предметной области исследования	16
1.3. Требования к информационной системе по сбору и обработке данных	17
Глава 2. Аналитический обзор существующих средств для обработки и визуализации данных.....	18
2.1. Системы для мониторинга социальных сетей.....	18
2.1.1. Медиалогия	18
2.1.2. Babkee	18
2.1.3. IQBuzz	19
2.2. Средства для визуализации данных.....	20
2.2.1. DataHero	20
2.2.2. Tableau	20
Глава 3. Проектирование информационной системы по обработке и визуализации данных из социальных сетей	22
3.1. Проектирование бизнес-процесса	22
3.2. Проектирование потоков данных	24
Глава 4. Разработка информационной системы по сбору и обработке данных из социальных сетей	25
4.1. Обоснование выбора программных средств разработки.....	25
4.2. Описание ролей, взаимодействующих с информационной системой	26
4.3. Описание основных алгоритмов по обработке данных	27
4.4. Демонстрация работы информационной системы.....	28
Глава 5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение ...	35
5.1. Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.....	36
5.1.1. Потенциальные потребители результатов исследования	36
5.1.2. Анализ конкурентных технических решений	37
5.1.3. Технология QuaD	38

5.1.4. SWOT-анализ	39
5.2. Планирование научно-исследовательских работ	40
5.2.1. Структура работ в рамках научного исследования.....	40
5.2.2. Определение трудоемкости выполнения работ	41
5.2.3. Разработка графика проведения научного исследования.....	44
5.2.4. Бюджет научно-технического исследования (НТИ).....	44
5.3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	48
Глава 6. Социальная ответственность.....	51
6.1. Правовые и организационные вопросы обеспечения безопасности	51
6.1.1. Организационные мероприятия обеспечения безопасности.....	51
6.1.2. Особенности законодательного регулирования проектных решений.....	53
6.2. Производственная безопасность	54
6.2.1. Анализ опасных и вредных производственных факторов.....	55
6.2.2. Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя (работающего)	56
6.3. Экологическая безопасность	57
6.3.1. Безопасность в чрезвычайных ситуациях.....	58
Вывод по разделу «Социальная ответственность»	60
Заключение	61
Список использованных источников	62
Приложение А. Листинг программы	63

Введение

За последние несколько лет произошел резкий скачок популярности социальных медиа. По данным сайта Statista количество пользователей социальных сетей с 2010 года до 2021 года вырастет с 0,97 до 3,02 миллиарда человек [1]. Сегодня в мире насчитывается 5,11 миллиарда уникальных мобильных пользователей, что на 100 миллионов (2 процента) больше, чем в прошлом году. В 2019 году 4,39 миллиарда пользователей Интернета выросли на 366 миллионов (9 процентов) по сравнению с январем 2018 года. В 2019 году в мире насчитывается 3,48 миллиарда пользователей социальных сетей, а по всему миру этот показатель вырос на 288 миллионов (9 процентов) с прошлого года. 3,26 миллиарда человек используют социальные сети на мобильных устройствах в январе 2019 года, при этом рост новых пользователей составил 297 миллионов, что на 10% больше, чем в годовом исчислении [2].

Технологии совершенствуются, пользователи решают свои проблемы посредством медийных платформ, начиная от продвижения своих брендов, заканчивая личным общением. Все это также означает увеличение объемов данных, которые можно анализировать в интересах бизнеса, маркетинга, социологии или политики. Так, например, в Facebook каждый час отправляется около 250 миллионов сообщений, а в Instagram за это же время пользователи оставляют примерно 100 миллионов лайков [3]. Для социальной сети Вконтакте статистика выглядит следующим образом: 1 миллиард отметок «Мне нравится» и 6,5 миллиардов сообщений в сутки.

Социальные медиа – интернет-приложение, позволяющее осуществлять передачу данных в оба конца, то есть потребитель может стать публикатором и наоборот. Понятие «пользовательский информационный материал» (с англ. User Generated Content) начало широко использоваться в 2005 году, когда возможности социальных сетей значительно выросли. UGC сближает бизнес и клиентов, дает возможность социально активным людям заявить о себе, а также упрощает процесс получения обратной связи брендам.

Можно выделить главные цели использования социальных медиа:

- Быть на связи с друзьями, семьей, коллегами
- Вести блог и узнавать новости
- Быть на связи с профессиональной сетью
- Делиться фото/видео/аудио
- Находить ответы на свои вопросы

Многие организации в разных сферах проводят анализ социальных сетей с целью улучшения своего бизнеса. Например, на основе анализа, интернет-магазин может автоматически предлагать товары, сопутствующие уже выбранному.

Маркетинг в социальных медиа (с англ. Social Media Marketing) активно проводят мониторинг позитивной и негативной информации о товаре или услуге в социальных сетях.

Исходя из личных убеждений, многие пользователи предпочитают скрывать информацию о себе, например, открыв доступ «только для друзей» или вовсе не заполнять свои данные. Несмотря на то, что большинство социальных медиа предлагает готовые API, открывающие доступ к данным, объем все также зависит от пользователя.

Даже при успешном сборе и обработке данных, исследование может быть бесполезным, если не будет установлена причинно-следственная связь. Прежде чем приступить к работе, необходимо сформулировать вопросы, на которые должно ответить исследование. Популярный случай интересных, но не объясненных совпадений был опубликован в одном из знаменитых медицинских журналов, который установил взаимосвязь между количеством шоколада, потребленного на душу населения, и числом Нобелевских лауреатов [5].

Анализ социальных сетей предполагает под собой стратегии исследования социальных структур, таких как дружеские связи или знакомства. Для эффективного представления и передачи результатов используется визуализация данных. Даже

самые сложные данные могут быть легко читаемыми благодаря хорошей визуализации.

Социальные сети сегодня являются основным источником данных, характеризующих общественные связи, процессы и отношение к событиям, людям и организациям. Применительно к высшим учебным заведениям, например, анализ социальных сетей позволит решить задачи, связанные с поиском и привлечением абитуриентов, в том числе определения сфер их интересов и ожиданий от учебы в университете. Помимо этого, анализ активностей пользователей, ассоциирующих себя с ВУЗом, позволит оценить репутацию университета в социальной сети, а также при необходимости запланировать корректирующие мероприятия.

Таким образом, задачи разработки и развития алгоритмов и программного обеспечения автоматизированного поиска и анализа данных социальных сетей являются актуальными.

Объект исследования: социальная сеть Вконтакте.

Предмет исследования: анализ и визуализация данных.

Цель работы: разработка алгоритмического и программного обеспечение анализа и визуализации данных социальных сетей.

Для достижения цели необходимо выполнить следующие задачи:

1. Постановка вопросов, на которые должно ответить исследование, и определение требований к продукту;
2. Аналитический обзор существующих методов обработки данных и визуализации;
3. Проектирование информационной системы по обработке и визуализации данных;
4. Разработка алгоритмического и программного обеспечения поиска, анализа и визуализации данных;

Глава 1. Исследование предметной области

1.1. Постановка задачи

С помощью алгоритмического и программного обеспечения поиска и анализа данных социальных сетей составить обобщенный портрет студента Томского Политехнического Университета. Собранная информация поможет ответить на следующие вопросы:

1. Процентное соотношение мужского и женского пола.
2. Насколько разнообразна география проживания студентов?
3. Возраст студентов.
4. Оставляют ли информацию о книгах, интересах и цитатах студенты в своих профилях?

Для наглядной демонстрации работы системы была составлена диаграмма IDEF0 (Рисунок 1.1)



Рисунок 1.1 – Модель работы программы

1.2. Описание предметной области исследования

Социальная сеть ВКонтакте – российская социальная сеть, позволяющая пользователям отправлять друг другу сообщения, создавать собственные страницы и сообщества, обмениваться медиа файлами и играть в браузерные игры. Сайт доступен на более, чем 90 языках.

На данный момент в данной социальной сети зарегистрировано более 460 миллионов пользователей. Из них 58869 человек отметили Томский Политехнический Университет как место учебы. В официальной группе университета «ТПУ | Томский Политехнический Университет» состоит 20521 человек, включая студентов, сотрудников и прочих.

1.3. Требования к информационной системе по сбору и обработке данных

Перед началом работы над программой, необходимо сформулировать функциональные и нефункциональные требования.

Функциональные требования:

1. Методы обработки и визуализации должны быть рациональными;
2. Визуализация результатов работы программы в графическом виде: диаграммы и социальный граф;
3. Собираемая информация должна быть деперсонализированной.

Нефункциональные требования:

1. Соответствие программы действующим законам и подзаконным актам РФ;
2. Программа должна работать без сбоев и аварийных завершений работы;
3. Программа должна быть разработана на языке программирования Python версии 3 в среде разработки PyCharm.

Глава 2. Аналитический обзор существующих средств для обработки и визуализации данных

2.1. Системы для мониторинга социальных сетей

2.1.1. Медиалогия

«Медиалогия» – автоматическая система мониторинга и анализа СМИ и социальных медиа.

В функции системы входят следующие пункты:

- Поиск по 800 миллионам источников, в числе которых «ВКонтакте», «Живой Журнал», «Телеграм», «Одноклассники», «Яндекс.Маркет» и многие другие ресурсы.
- Автоматическое определение тональности сообщений, наиболее интересных инфоповодов и влиятельных авторов.
- Антиспам-модуль, который исключает нерелевантные сообщения, и лингвистический модуль, обеспечивающий нахождение всех упоминаний бренда.
- Сбор дополнительной информации по аудитории, вовлеченности и другим важным маркетинговым и PR-метрикам.
- Визуализация данных и удобная выгрузка отчетов (в том числе можно экспортировать ID авторов, например, для настройки таргетированной рекламы).

Тариф «Есопому» для малого бизнеса – стоит 20 000 рублей в месяц. Для агентств подписка будет стоить дороже на 20 %. Есть бесплатный пробный доступ 7 дней.

2.1.2. Babkee

«Babkee» – российский сервис мониторинга упоминаний и анализа соцмедиа, управления репутацией и изучения активности конкурентов.

Функциональность:

- Сбор данных по 46 миллионам площадок (в их числе тематические порталы, соцсети, видеохостинги, СМИ и др.).
- Смысловой анализ текста и его тональности.
- Выявление первоисточников и дублей, определение значимости сообщения.
- Подробная отчетность по авторам упоминаний (возраст, регион, контакты и др.).
- Коллективная обработка сообщений, система уведомлений.
- Экспорт в Excel.

Минимальная стоимость месячной подписки на сервис – 8 500 рублей. На VA Express действует бесплатный тестовый период, он длится 7 дней.

2.1.3. IQBuzz

«IQBuzz» – система мониторинга социальных сетей, блогов, форумов, СМИ. Можно использовать для управления репутацией, анализа конкурентов и эффективности рекламных кампаний, исследования целевой аудитории.

Функции:

- Поиск отзывов, трендов, лидеров мнений по 10 000+ источникам, включая LiveJournal, VK, YouTube, Instagram, Twitter.
- Автоматическое определение тональности и ее персональная настройка.
- Сбор статистики (пол, возраст аудитории и др.), отслеживание KPI (охват, вовлеченность и др.).
- Выгрузка отчетов в Excel и Word.
- Командная работа с разграничением прав доступа.
- Простановка тегов по задачам, авторам, документам.
- Интеграция с CRM, возможность работать через API.
- Бесплатная ретроспектива до 10 лет.

Тарифы на IQBuzz стоят: от 7 900 до 54 900 рублей в месяц. Есть бесплатный пробный период 7 дней.

2.2. Средства для визуализации данных

2.2.1. DataHero

DataHero – полезный инструмент для визуализации данных, который можно использовать для построения графиков и диаграмм.

Функционал:

- Соединение с данными (данные могут храниться и в облачных хранилищах, и на компьютере пользователя);
- Комбинирование данных;
- Аналитика;
- Экспорт результатов.



Рисунок 2.1. – Визуализация с помощью DataHero

2.2.2. Tableau

Tableau – веб-сервис, позволяющий быстро и эффективно работать с данными, создавая графики. Каждому пользователю предоставляется до 10 Гб облачного пространства.

Возможности:

- Загрузка информации в систему и отслеживание динамики в реальном времени;

- Поддержка импорта из MS Excel, MS Access, различных текстовых форматов и серверных баз данных;
- Упрощенная сортировка, изменение видов диаграмм, рекомендации по визуализации;
- Интерактивные графики;
- Возможность публикации результатов в интернете;

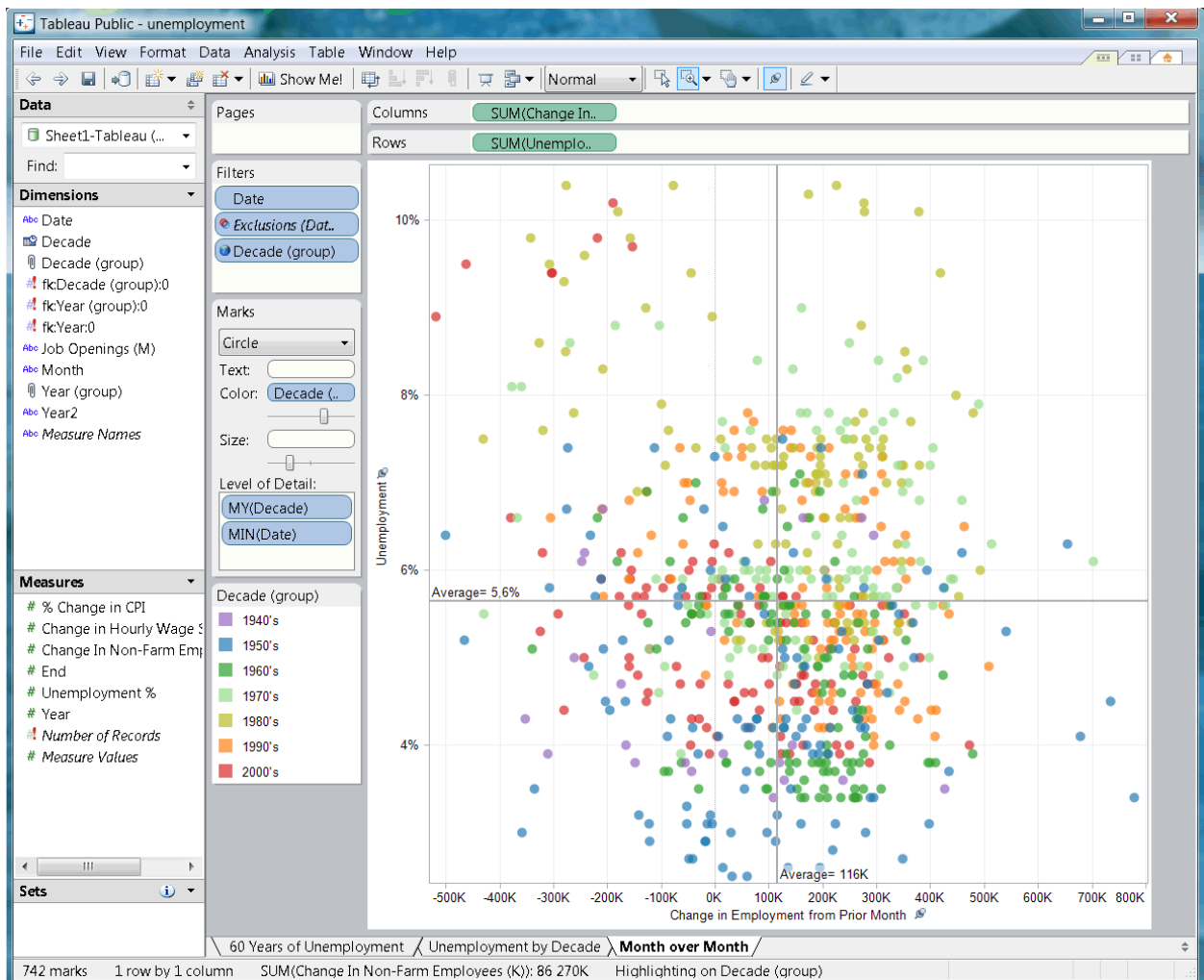


Рисунок 4.2 – Визуализация с помощью Tableau

Глава 3. Проектирование информационной системы по обработке и визуализации данных из социальных сетей

При разработке любого программного продукта важным этапом является проектирование.

Проектирование – это поиск способа, удовлетворяющего требованиям функциональности системы имеющимися средствами с учетом заданных ограничений.

3.1. Проектирование бизнес-процесса

На рисунке 3.1 представлена диаграмма EPC, демонстрирующая бизнес-процесс «Обработка и визуализация данных, собранных с социальной сети».

Событийная цепочка процессов – тип блок-схемы, используемой для бизнес-моделирования.

На данной диаграмме показаны все этапы обработки и визуализации данных, начиная с постановки цели, заканчивая формированием отчета.

Важным этапом является верификация данных. В файле с собранными данными у удаленных пользователей атрибут «deactivate» имеет значение «banned». Такие строки необходимо удалить. Также необходимо удалить пользователей, у которых в «education» и в «universities» не указан Томский Политехнический Университет. Таким образом, происходит исключение из исследования неподходящих пользователей.

В пункт «Обработать данные» входит подсчет количества пользователей, удовлетворяющих определенным критериям:

- 1) Пол;
- 2) География проживания (страна, город);
- 3) Семейное положение (свободен/в отношениях);
- 4) Интересы (книги, цитаты).

Визуализация результатов включает в себя построение круговых диаграмм.



Рисунок 3.1. – Диаграмма EPC

3.2. Проектирование потоков данных

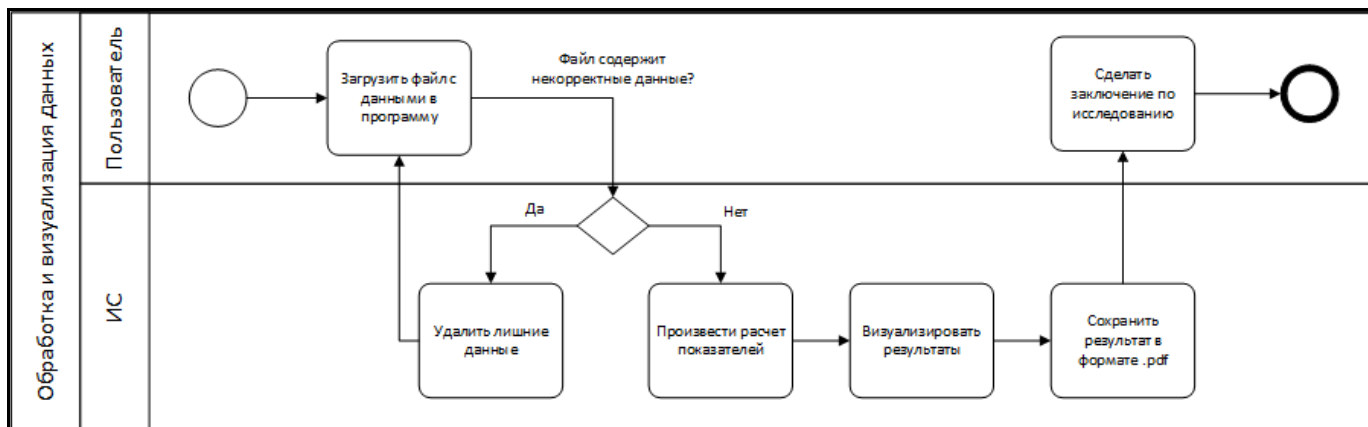


Рисунок 3.2 – Диаграмма BPMN процесса обработки и визуализации данных

BPMN (The Business Process Modeling/Management Notation) – система условных обозначений для построения схемы протекания бизнес-процессов. Данная диаграмма демонстрирует взаимодействие пользователя с информационной системой.

На рисунке 3.2 показано, что пользователь начинает и заканчивает работу информационной системы после получения всех необходимых результатов.

Процесс проведения исследования практически полностью автоматизирован, за исключением формирования итогового отчета. Несмотря на весь технический прогресс, с данной задачей человек справится лучше.

Глава 4. Разработка информационной системы по сбору и обработке данных из социальных сетей

4.1. Обоснование выбора программных средств разработки

Для наиболее рациональной и эффективной работы программы, необходимо подобрать соответствующие программные средства разработки.

Во-первых, необходимо подключить библиотеку для работы с файлами excel. Для языка программирования Python существует два способа взаимодействия: через библиотеки и использованием com-объектов.

- xlrd

Принцип работы: получение объекта книги. Есть возможность брать отдельные листы книги, не сохраняя всю книгу в память.

- openpyxl

Данная библиотека, в отличии от предыдущей, позволяет не только считывать, но и записывать данные в excel файл. Один из минусов: нет возможности использовать формат excel с.

Openpyxl дает возможность обращаться к ячейкам и последовательность через их имена, что довольно удобно и понятно при чтении кода.

- win32com (из пакета pywintypes)

При работе с win32com файл excel используется через com-объект, что позволяет выполнять с файлом все операции, которые позволяет делать Excel с использованием VBA.

Таблица 4.1 – Морфологический анализ библиотек для работы с excel

Метрика	Вес метрики	xlrd, xlwt	openpyxl	win32com
Совместимость с форматами	0,5	5	2	5
Наличие документации	0,3	4	4	4
Простота использования	0,1	5	3	2
Опыт работы с библиотекой	0,1	5	2	1
Итого:	1	4,7	2,7	4,2

Исходя из таблицы 1, приходим к выводу, что наиболее подходящим вариантом является использование библиотек xlrd, xlwt.

Во-вторых, выбранные данные необходимо визуализировать. Исследование подразумевает представление информации в виде круговых диаграмм.

Matplotlib – библиотека на языке программирования Python для визуализации данных двумерной графикой. Получаемые изображения могут быть использованы в качестве иллюстраций в публикациях.

Пакет plotly – графическая библиотека, позволяющая строить графики и диаграммы. Plotly – компания, производящая несколько продуктов и инструментов с открытым исходным кодом. Библиотека бесплатна для использования и позволяет создавать неограниченное количество графиков в автономном режиме.

Таблица 4.2 – Морфологический анализ средств визуализации данных

Метрика	Вес	Matplotlib	plotly
Возможность создания диаграмм других типов	0,2	4	5
Работа офлайн	0,1	5	1
Наличие бесплатного доступа	0,2	5	5
Простота использования	0,3	4	5
Наличие обучающего материала	0,2	5	4
Итого:	1	4,5	4,4

Таким образом, визуализация данных в виде круговых диаграмм будет производиться с помощью Matplotlib.

4.2. Описание ролей, взаимодействующих с информационной системой

Разрабатываемая система предназначена для проведения исследований, то есть пользователями являются специалисты в области маркетинга, политики и так далее.

На рисунке 4.1 приведена диаграмма вариантов использования программы, на которой показаны основные действия, совершаемые с программой.

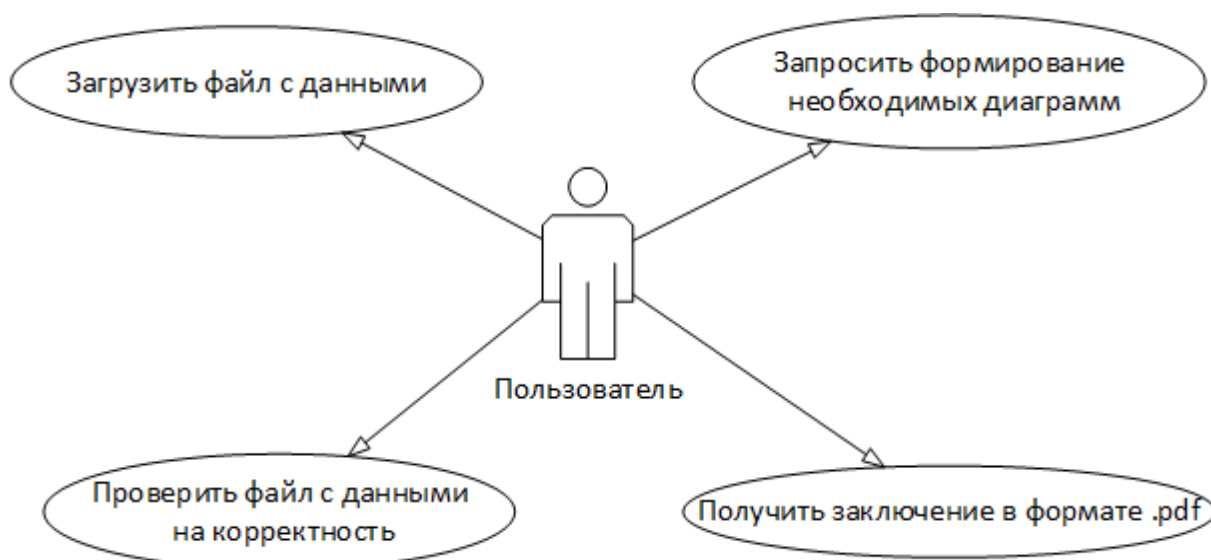


Рисунок 4.1 – Диаграмма вариантов использования

4.3. Описание основных алгоритмов по обработке данных

Файл с собранными данными содержал большое количество ненужной информацией: пустые поля, информация об удаленных или заблокированных пользователях и неполная информация (дата рождения без года).

Для решения данной проблемы была написана программа на VBA, сортирующая и удаляющая лишние данные.

VBA (Visual Basic for Applications) — это диалект языка Visual Basic, расширяющий возможности Visual Basic и предназначенный для работы с приложениями Microsoft Office и другими приложениями от Microsoft и третьих фирм.

Код программы представлен ниже:

```

Sub Del_SubStr()
  Dim sSubStr As String
  Dim lCol As Long
  Dim lLastRow As Long, li As Long
  Dim lMet As Long
  Dim arr

  sSubStr = InputBox("Укажите значение, которое необходимо найти в строке",
  "Запрос параметра", "")
  If sSubStr = "" Then lMet = 0 Else lMet = 1
  lCol = Val(InputBox("Укажите номер столбца, в котором искать указанное
  значение", "Запрос параметра", 1))
  If lCol = 0 Then Exit Sub
  
```

```

lLastRow = ActiveSheet.UsedRange.Row - 1 +
ActiveSheet.UsedRange.Rows.Count
arr = Cells(1, lCol).Resize(lLastRow).Value
Application.ScreenUpdating = 0
Dim rr As Range
For li = 1 To lLastRow 'цикл с первой строки до конца
  If -(InStr(arr(li, 1), sSubStr) > 0) = lMet Then
    If rr Is Nothing Then
      Set rr = Cells(li, 1)
    Else
      Set rr = Union(rr, Cells(li, 1))
    End If
  End If
Next li
If Not rr Is Nothing Then rr.EntireRow.Delete
Application.ScreenUpdating = 1
End Sub

```

После работы алгоритма из 20 606 строк данных осталось 18 426.

В приложении А представлен код для построения круговых диаграмм с использованием библиотеки визуализации Matplotlib.

4.4. Демонстрация работы информационной системы

Результатом выполнения данной работы являются диаграммы, демонстрирующие процентное соотношение по тому или иному признаку.

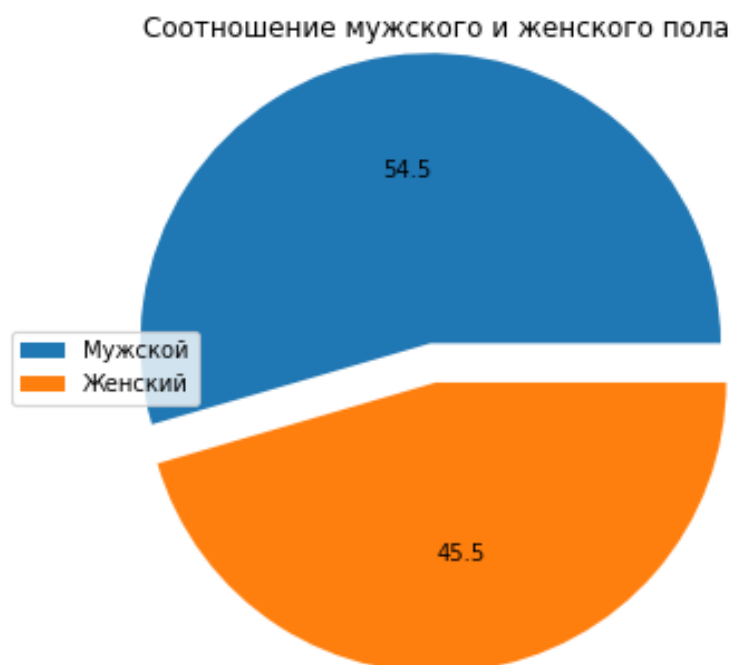


Рисунок 4.2 – Процентное соотношение лиц мужского и женского пола

На данной диаграмме видно, что лиц мужского пола на 10% больше, чем женского. Этот факт опровергает стереотип о том, что в техническом ВУЗе обучаются преимущественно парни.

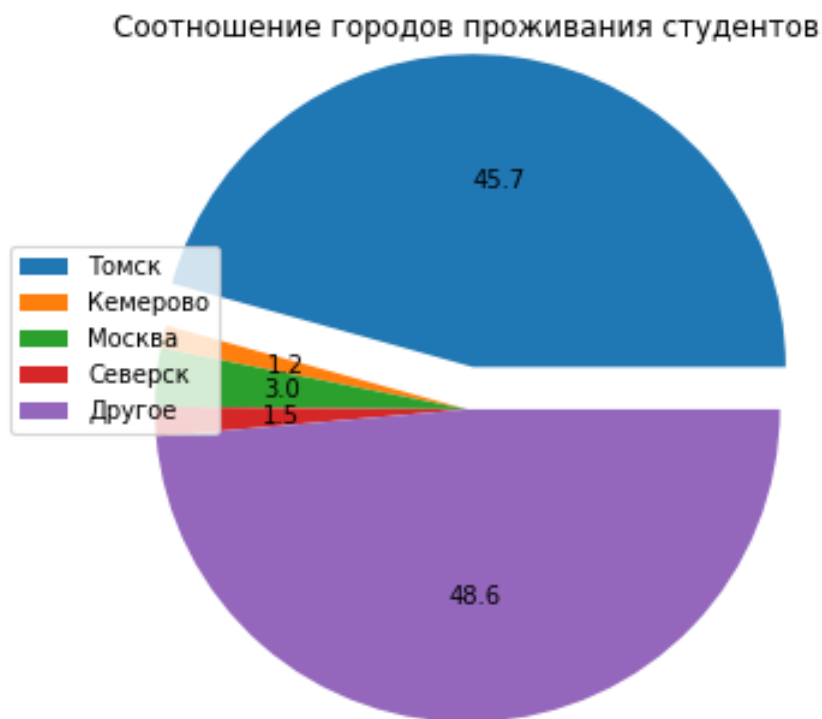


Рисунок 4.3 – Количественное соотношение по городам

Как показывает диаграмма, в Томском Политехническом Университете больше иногородних студентов, чем жителей Томска. Это говорит о хорошей агитационной кампании. Помимо дней открытых дверей и выездных мероприятий, для абитуриентов разработана игра JOIN.TPU, где каждый игрок — студент, только что приехавший покорять Томск. Участник проходит ролевые этапы видео-квеста, которые помогают ему собирать необходимые для победы бонусы и попутно решать задачи по физике, математике и другим дисциплинам. Игрок «живет» и «учится» в Томском политехе: посещает лекции, выполняет лабораторные работы, проводит студенческий досуг, занимается общественной деятельностью, волонтерской работой [6].

Соотношение стран проживания студентов

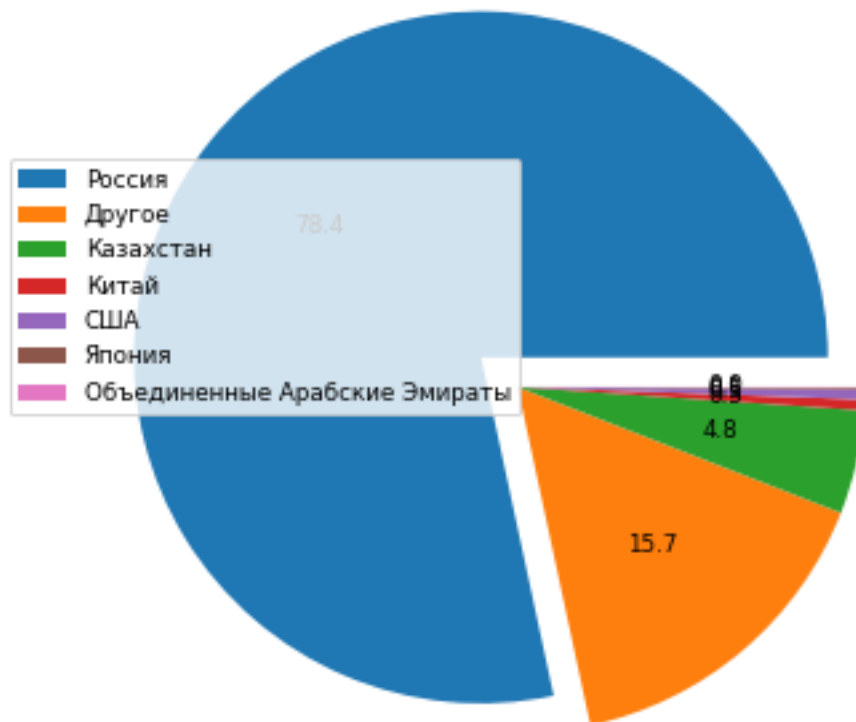


Рисунок 4.4 – Количественное соотношение по странам

Диаграмма на рисунке 4.4 показывает, что Россия занимает первое место по количеству студентов, обучающихся в ТПУ. Также большое количество студентов из Казахстана, Китая, присутствуют даже студенты из Объединенных Арабских Эмиратов и Соединенных Штатов Америки. Опять же можно отметить эффективность игры JOIN.TPU. Как отмечают организаторы, самыми активными игроками стали абитуриенты из разных уголков России, Украины, Казахстана, Беларуси, Киргизии, Армении и Узбекистана. Присоединились к игре в студенчество и пользователи из США, Германии, Италии, Польши, Латвии, Молдовы, Китая [6].

Соотношение студентов по семейному положению

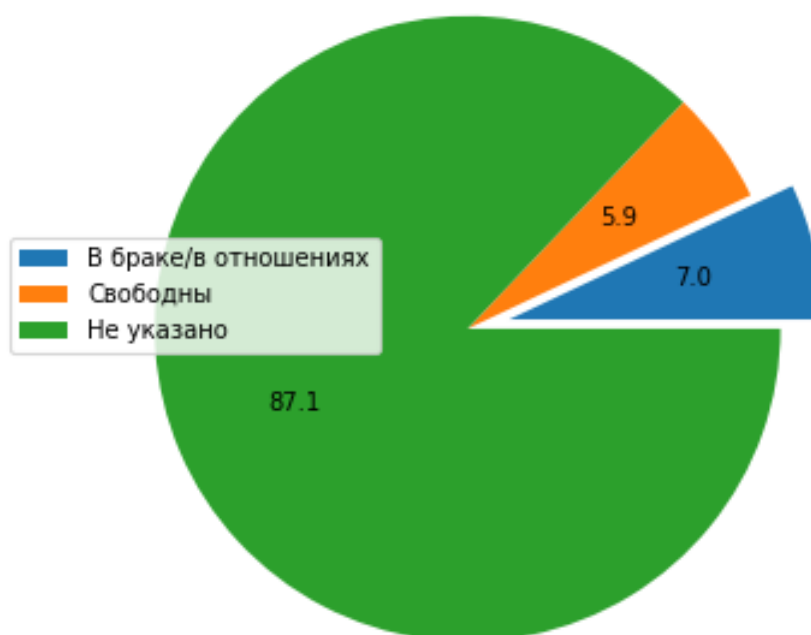


Рисунок 4.5 – Процентное соотношение по семейному положению

Большое количество статуса «не указано» оправдано тем, что люди предпочитают скрывать свою личную жизнь в социальных сетях. Тем не менее, около 1000 человек указали, что они состоят в браке, или имеют отношения.

По результатам исследования, проведенного Jobvite, 93% интервьюеров предпочитают просмотреть профили собеседующегося в наиболее крупных социальных сетях. 44% интервьюеров отмечают, что такой способ поиска сотрудников повышает качество их работы, 44% замечают рост числа нанятых и 34% говорят, что использование социальных сетей позволило сократить время, нужное на закрытие одной вакансии [7]. То, что не было указано в резюме, может быть отмечено на странице в социальной сети и наоборот. В конечном счете, если работодателю будет нужна какая-либо информация, он ее найдет.

Большую роль при трудоустройстве имеет наличие или отсутствие семьи. Так, для работы с ненормированным графиком и частыми командировками, подходит

человек, не находящийся в браке. Наличие детей также создает некоторые сложности. Тем не менее, семейный человек вызывает чувство стабильности и надежности.

На рисунках 4.6-4.8 изображены диаграммы, демонстрирующие наличие или отсутствие информации о книгах, цитатах и интересах. Результаты каждого исследования показывают, что лишь 3,7-5,1% студентов заполняют эти поля на своей странице в социальной сети.

В наше время термин «Цифровой след» приобрел широкое распространение в связи с увеличением информации в интернете и развитием технологий, позволяющих сделать это любым удобным способом.

Цифровой след – совокупность данных, которые пользователь генерирует во время пребывания в Интернете. Существует активный и пассивный цифровой след. Главное различие между ними заключается в том, что одни данные пользователь оставляет намеренно (посты в блог, комментарии и так далее), а другие собираются автоматически (ip-адреса устройств, истории посещения и так далее).

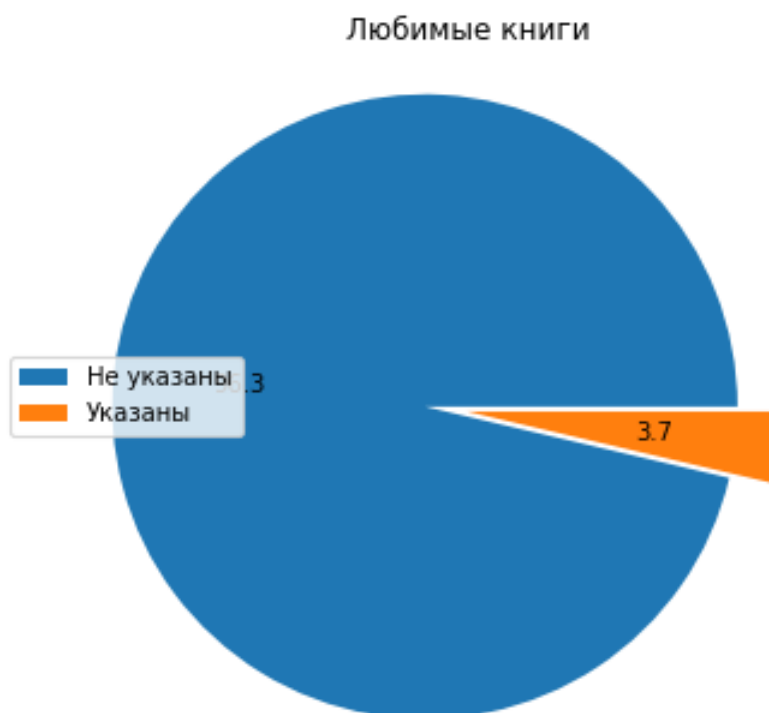


Рисунок 4.6 – Процентное соотношение по книгам

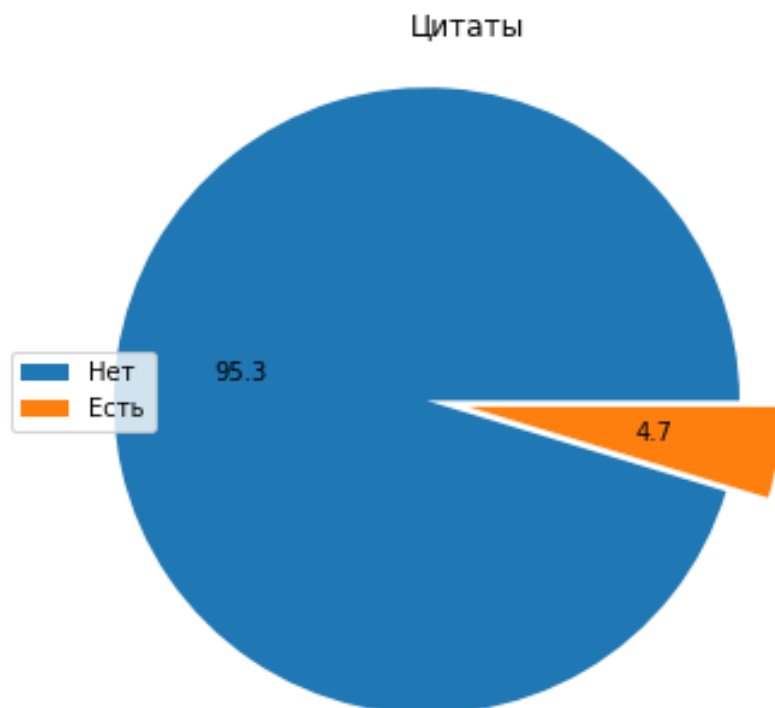


Рисунок 4.7 – Процентное соотношение по цитатам

Чаще всего, в чистоте или даже отсутствие своего цифрового следа, заинтересованы диссиденты, шпионы, преступники и журналисты. Для большинства обычных пользователей цифровой отпечаток не несет никакой опасности, но тем не менее, сотни тысяч людей по всему миру желают стереть свою цифровую тень.

Это обосновывается некоторыми причинами. Например, злоумышленники могут получить доступ к личным аккаунтам и даже к банковским счетам.

С появлением Интернета и социальных сетей в частности, цифровой след стал источником информации для создания и рекомендации рекламы пользователю, который эту информацию оставил.

Content Marketing Institute совместно с MarketingProfs провели исследование на тему эффективности рекламы в Интернете.

Рекламные посты (к примеру, публикации в Facebook или Twitter и Pinterest) применяют 76% B2C маркетологов. 61% из них подтвердили действенность этого вида рекламы в социальных сетях, эффективность которого была оценена на 4 или 5 баллов по пятибалльной системе (3 демонстрирует нейтральное отношение).



Рисунок 4.8 – Процентное соотношение по интересам

Таким образом, не заполняя поля «книги», «цитаты», «интересы», пользователь защищает себя от ненужной и навязчивой рекламы.

Помимо этого, нежелание делиться личной информацией оправдано совмещением нишевых сообществ. Наличие в списке друзей членов семьи или коллег вызывает определенные психологические проблемы, так как человек многогранен, и каждая грань соприкасается с разными сообществами.

Глава 5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Осуществляя любую научно-исследовательскую работу, необходимо провести ее экономическое обоснование. Экономическое обоснование включает в себя: определение потенциальных потребителей и сегмента рынка, сравнительный анализ с конкурентами, оценка себестоимости и определение трудоемкости выполнения работ. Это необходимо при поиске источников финансирования и оценке коммерческой ценности проекта.

Данный раздел помогает понять, будет ли продукт востребованным на рынке, сколько он будет стоить, какой срок понадобится для его реализации и так далее. Цель раздела: спроектировать и создать конкурентоспособные технологии, отвечающие современным требованиям в области ресурсоэффективности и ресурсосбережения.

Научно-исследовательская работа представляет собой алгоритмическое и программное обеспечение поиска и анализа данных социальных сетей для составления обобщенного портрета студента Томского Политехнического Университета. Программный продукт позволяет установить не всегда очевидные закономерности между данными в социальных сетях и некоторыми факторами. В рамках предлагаемой работы будут спроектированы и реализованы комплекс алгоритмов и программных модулей для автоматизированного параметризованного поиска, извлечения, хранения и анализа открытой информации социальных сетей. Прикладное назначение разрабатываемой системы – информационная поддержка маркетинговых, социологических и политологических исследований. Приоритетным источником данных является социальная сеть ВКонтакте.

5.1. Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

5.1.1. Потенциальные потребители результатов исследования

Для анализа потребителей необходимо рассмотреть целевой рынок и сегментировать его.

Целевой рынок – группа людей, которым планируется продать товар или услугу.

Сегментация рынка – разбивка рынка на сегменты по различным признакам. Существуют географические, демографические, поведенческие и другие критерии, также возможно их комбинирование с некоторыми характеристиками, такими как, пол, возраст, национальность и прочее.

Предполагается, что пользователями разработки будут являться ученые различных областей. Целесообразно использовать два наиболее значимых критерия: размер компании-заказчика и область исследования.

Таблица 1 – Сегментирование рынка

		Область исследования				
		Политика	Социология		Маркетинг	Экономика
Размер компании	Крупные					
	Средние					
	Мелкие					

Brand Analytics	ВЦИОМ	Фонд «Социум»

Как видно из таблицы 1, компании, занимающиеся исследованиями по данным из социальных сетей, заняли практически все позиции среднего звена. На начальном этапе стоит ориентироваться на мелкие компании в областях исследования: политика, социология и маркетинг.

Однако, с развитием социальных сетей возрастает спрос на различные исследования у государства, а это значит, что стоит учесть расширение сегмента на крупных заказчиков.

5.1.2. Анализ конкурентных технических решений

Для того, чтобы проект был конкурентоспособным, необходимо периодически проводить детальный анализ конкурирующих разработок. Важно правильно оценить все слабые и сильные стороны.

Целесообразно проводить данный анализ с помощью оценочной карты, которая представлена в таблице 2.

Таблица 2 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
Интуитивно-понятный интерфейс	0,1	5	2	4	0,5	0,2	0,4
Кроссплатформенность	0,05	3	1	4	0,15	0,05	0,2
Сопроводительная документация	0,01	5	2	5	0,05	0,02	0,05
Простота ввода в эксплуатацию	0,1	4	4	4	0,4	0,4	0,4
Рациональные методы обработки данных	0,1	4	2	3	0,4	0,2	0,3
Корректная визуализация результатов	0,1	5	5	5	0,5	0,5	0,5
Язык написания программы	0,1	3	4	2	0,3	0,4	0,2
Экономические критерии оценки эффективности							
Цена	0,24	4	5	2	0,96	1,2	0,48
Предполагаемый срок эксплуатации	0,1	4	4	4	0,4	0,4	0,4

Конкурентоспособность продукта	0,1	5	5	5	0,5	0,5	0,5
Итого	1				4,16	3,87	3,43

Вывод: анализ конкурентоспособности показал, что технические и экономические показатели заметно опережают показатели конкурентов. Важными преимуществами стали интуитивно-понятный интерфейс и рациональные методы обработки данных, а также важным аспектом является относительно невысокая цена.

5.1.3. Технология QuaD

Технология QuaD (QUality ADvisor) представляет собой гибкий инструмент измерения характеристик, описывающих качество новой разработки и ее перспективность на рынке и позволяющие принимать решение целесообразности вложения денежных средств в НИР.

Таблица 3 – Оценочная карта для сравнения конкурентных технических решений

Критерий оценки	Вес критерия	Баллы	Макс. балл	Относительное значение (3/4)	Средневзвешенное значение (5*2)
1	2	3	4	5	6
Показатели оценки качества разработки					
Интуитивно-понятный интерфейс	0,1	80	100	0,8	0,08
Кроссплатформенность	0,05	50	100	0,5	0,025
Сопроводительная документация	0,01	90	100	0,9	0,009
Простота ввода в эксплуатацию	0,1	60	100	0,6	0,06
Рациональные методы обработки данных	0,1	85	100	0,85	0,085
Корректная визуализация результатов	0,1	90	100	0,9	0,09
Язык написания программы	0,1	60	100	0,6	0,06
Показатели оценки коммерческого потенциала разработки					
Цена	0,24	70	100	0,7	0,168

Предполагаемый срок эксплуатации	0,1	80	100	0,8	0,08
Конкурентоспособность продукта	0,1	90	100	0,9	0,09
Итого	1			7,55	0,747

Средневзвешенное значение показателя качества и перспективности научной разработки составляет 74,7, что говорит о том, что перспективность выше среднего.

5.1.4. SWOT-анализ

SWOT анализ – инструмент стратегического менеджмента, представляющий собой комплексный анализ научно-исследовательского проекта. SWOT анализ применяют для исследования внешней и внутренней среды проекта.

Таблица 4 – Матрица SWOT

	<p>Сильные стороны научно-исследовательского проекта:</p> <p>С1. Заявленная экономичность и энергоэффективность технологии.</p> <p>С2. Экологичность технологии</p> <p>С3. Более низкая стоимость по сравнению с аналогами</p> <p>С4. Простота использования</p> <p>С5. Быстрая обработка данных благодаря рациональным формулам</p>	<p>Слабые стороны научно-исследовательского проекта:</p> <p>Сл1. Сложность разработки</p> <p>Сл2. Необходимость постоянной поддержки и обновления ПО после внедрения</p> <p>Сл3. Длительное ожидание обработки запросов на сбор данных (в частности у Twitter)</p> <p>Сл4. Отсутствие рекламной кампании</p> <p>Сл5. Невозможность получения достоверной картины ввиду не полных данных, введенных пользователем в социальной сети.</p>
<p>Возможности:</p> <p>В1. Появление спроса на исследования других видов.</p> <p>В2. Расширение области применения технологии</p> <p>В3. Использование полученных обработанных данных для агитационной кампании университета</p>	<p>Благодаря удобному интерфейсу, быстрой обработке данных и низкой стоимости, данной разработкой могут заинтересоваться не только другие ВУЗы, но и компании, и даже государственные учреждения, которые могут</p>	<p>Несмотря на все сильные стороны, технология все же не так популярна, как ее конкуренты. Решением может стать повышение стоимости конкурентных разработок, при котором выбор заказчика будет склоняться к наиболее</p>

В4.Повышение стоимости конкурентных разработок	проводить различные исследования, решающие их проблемы, по полученным данным из научной разработки.	выгодному товару, каким и является наш продукт.
Угрозы: У1. Развитая конкуренция У2. Снижения популярности и/или необходимости проведения исследований. У3. Запрет на использование данных из социальных сетей У4. Прекращение существования социальной сети ВКонтакте	Сильные стороны научной разработки решают первую угрозу, но на остальные со стороны разработчиков повлиять невозможно. При У4 имеет смысл обрабатывать данные из другой социальной сети.	Основной угрозой является развитая конкуренция. Высока вероятность, что заказчики могут предпочесть встроенные инструменты обработки данных, если такие появятся, вместо использования каких-либо сторонних приложений.

По результатам SWOT-анализа были выявлены сильные и слабые стороны научной разработки, а также ее угрозы и возможности. Некоторые слабые стороны можно компенсировать возможностями, а угрозы – сильными сторонами.

5.2. Планирование научно-исследовательских работ

5.2.1. Структура работ в рамках научного исследования

Для планирования комплекса предполагаемых работ необходимо выполнить следующие задачи:

1. Определить структуры работ в рамках научного исследования
2. Определить участников каждой работы
3. Установить продолжительность работ
4. Построить график проведения научных исследований

Таблица 5 – Перечень этапов, работ и распределение исполнителей

Основные этапы	№ раб	Содержание работы	Должность исполнителя
Разработка технического задания	1	Составление и утверждение технического задание	Руководитель проекта

Проектирование разработки	2	Определение целей исследования	Руководитель проекта
	3	Составление календарного плана	Лаборант
	4	Составление диаграмм	Лаборант
Разработка приложения	5	Определение средств разработки	Лаборант
	6	Анализ подходящих алгоритмов для обработки результатов	Лаборант
	7	Разработка макетов приложения	Лаборант
	8	Программирование, отладка приложения	Лаборант
	9	Тестирование	Лаборант
Внедрение приложения	10	Развертывание ИС в сети	Лаборант, Руководитель
	11	Оформление и утверждение акта сдачи-приемки системы	Руководитель
Оформление отчета по ВКР	12	Составление пояснительной записки	Лаборант

5.2.2. Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Для определения ожидаемого значения трудоемкости используется следующая формула:

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

где $t_{ож\ i}$ – ожидаемая трудоемкость выполнения i -ой работы чел.-дн. ;

$t_{min\ i}$ – минимально возможная трудоемкость выполнения заданной i -ой работы, чел.-дн.;

$t_{max\ i}$ – максимально возможная трудоемкость выполнения заданной i -ой работы

Исходя из этого, определяется продолжительность каждой работы в рабочих днях T_{pi} , учитывающая параллельность выполнения работ несколькими исполнителями.

$$T_{pi} = \frac{t_{ож\ i}}{Ч_i},$$

где T_{pi} – продолжительность одной работы, раб. дн;

$t_{ож\ i}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.;

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Таблица 6 – Временные показатели проведения научного исследования

Название работы	Трудоемкость									Длительность работ в рабочих днях T_{pi}	Длительность работ в календарных днях T_{ki}				
	t_{min}			t_{max}			$t_{ожид}$				Исп. 1	Исп. 2	Исп. 3		
	Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3						
1. Составление и утверждение технического задания (Руководитель)	3	2	4	6	7	5	4	4	4	4	4	4	5	5	5
2. Определение целей исследования (Руководитель)	2	3	2	5	5	4	3	4	3	3	4	3	4	5	4
3. Составление календарного плана (Лаборант)	1	2	1	4	5	3	2	3	2	2	3	2	2	4	2
4. Составление диаграмм (Лаборант)	5	6	4	7	8	10	6	7	6	6	7	6	7	9	7

5.Определение средств разработки (Лаборант)	1	2	1	3	5	4	2	3	2	2	3	2	2	4	2
6.Анализ подходящих алгоритмов для обработки результатов (Лаборант)	3	2	4	5	4	6	4	3	5	4	3	5	5	4	6
7.Разработка макетов приложения (Лаборант)	5	4	6	8	7	7	6	5	6	6	5	6	7	6	7
8.Программирование, отладка приложения (Лаборант)	14	10	11	17	15	16	15	12	13	15	12	13	18	15	16
9.Тестирование (Лаборант)	5	4	6	7	8	7	6	6	6	6	6	6	7	7	7
10.Развертывание ИС в сети (Руководитель, Лаборант)	5	7	6	3	4	4	4	6	5	2	3	3	2	4	3
11.Оформление и утверждение акта сдачи-приемки системы (Руководитель)	1	2	2	4	4	5	2	3	3		3	3	2	4	4
12.Составление пояснительной записки (Лаборант)	2	1	2	4	4	5	3	2	3	3	2	3	4	2	4

Таблица 7 – Продолжительности работ для исполнителей

	Длительность работ в рабочих днях		Длительность работ в календарных днях	
	Руководитель	Лаборант	Руководитель	Лаборант
Исполнитель 1	11	46	13	54
Исполнитель 2	14	44	18	55
Исполнитель 3	13	46	16	54

5.2.3. Разработка графика проведения научного исследования

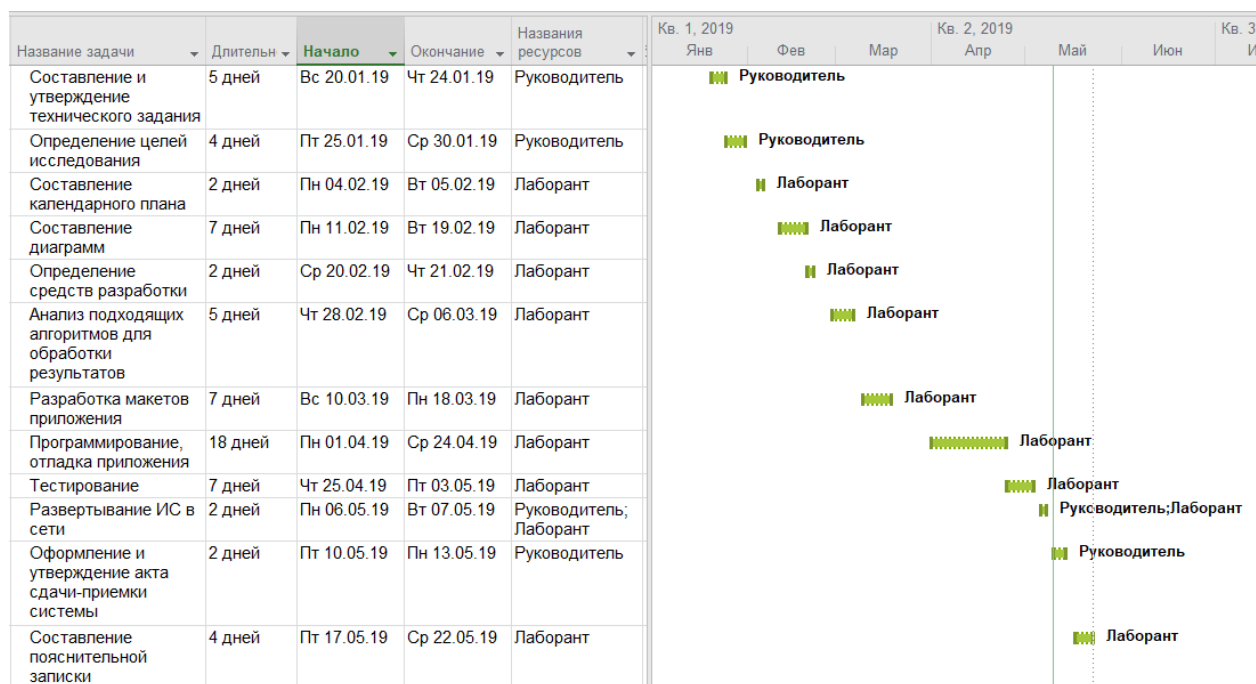


Рисунок 1 – Диаграмма Ганта

5.2.4. Бюджет научно-технического исследования (НТИ)

Расчет затрат на специальное оборудование для научных (экспериментальных) работ

В процессе работы над проектом планируется использование ноутбука, купленного заранее, поэтому следует рассчитать его амортизацию. Срок полезного использования составляет 3 года. Время написания ВКР – 5 месяцев. Тогда норма амортизации:

$$A_n = \frac{1}{n} * 100\% = \frac{1}{3} * 100\% = 33,33\%$$

Годовые амортизационные отчисления:

$$A_{\Gamma} = 30000 * 0,33 = 9900 \text{ рублей}$$

Ежемесячные амортизационные отчисления:

$$A_{\text{м}} = \frac{A_{\Gamma}}{12} = \frac{9900}{12} = 825 \text{ рублей}$$

Таблица 8 – Расчет бюджета затрат на приобретение спецоборудования для научных работ

№	Наименование	Кол-во единиц	Цена единицы, руб.	Общая стоимость, тыс. руб.	Амортизационные отчисления
1	Ноутбук Lenovo G50-45	1	30 000	30000	4125
2	ПО Microsoft office	1	4 684	4 684	644
Итого:				34 686	4769

Основная заработная плата исполнителей темы

Таблица 9 – Баланс рабочего времени

Показатели рабочего времени	Студент	Руководитель
Календарные дни	365	
Нерабочие дни (праздники/выходные)	66	
Потери рабочего времени		
-отпуск	56	56
-невыходы по болезни	4	2
Действительный годовой фонд рабочего времени	239	241

Затраты на заработную плату рассчитываются по следующей формуле:

$$Z_{\text{п}} = Z_{\text{осн}} + Z_{\text{доп}}, \text{ где}$$

$Z_{\text{доп}}$ – дополнительная заработная плата, руб;

$Z_{\text{осн}}$ – основная заработная плата, руб.

$$Z_{осн} = Z_{дн} * T_p * (1 + K_{пр} + K_{д}) * K_{р}, \text{ где}$$

$Z_{дн}$ – среднедневная заработная плата, руб.;

$K_{пр}$ – премиальный коэффициент (0,3);

$K_{д}$ – коэффициент доплат и надбавок (0,3-0,5);

$K_{р}$ – районный коэффициент (для Томска 1,3)

T_p – продолжительность работ, выполняемых работником, раб. дни

Среднедневная заработная плата:

$$Z_{дн} = \frac{Z_m * M}{F_d}$$

Z_m – месячный оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года (для 6-дневной рабочей недели $M=10,4$);

F_d – действительный годовой фонд рабочего времени персонала, раб.дн.

Таблица 10 – Расчет основной заработной платы

Исполнители	$Z_{дн}$, руб	$K_{пр}$	$K_{д}$	$K_{р}$	T_p	$Z_{осн}$
Студент	946,8	0,3	0,3	1,3	46	90 597
Научный руководитель	1452,7	0,3	0,4	1,3	11	35 315
Итого:						125 913

Дополнительная заработная плата исполнителей темы

Таблица 11 – Расчет дополнительной заработной платы

Исполнитель	Основная заработная плата, руб.	Коэффициент дополнительной заработной платы	Дополнительная заработная плата, руб
Студент	90 597	0,15	13 589,6
Руководитель	35 315		5 297,3
Итого:			18 886,95

Отчисления во внебюджетные фонды (страховые отчисления)

Таблица 12 – Расчет отчислений во внебюджетные фонды

Исполнитель	Основная заработная плата + дополнительная, руб.	Коэффициент отчислений во внебюджетные фонды	Сумма отчислений
Студент	104 187	0,28	29 172
Руководитель	40 612,98		11 371
Итого:			40 543

Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопия материалов, оплата услуг связи, электроэнергии и т.д.

$$\text{Знакл} = (\text{сумма статей 1-4}) * \text{Кнр}, \text{ где}$$

Кнр – коэффициент, учитывающий накладные расходы (16%)

Формирование бюджета затрат научно-исследовательского проекта

Таблица 13 – Бюджет затрат НТИ

Наименование статьи	Сумма, руб	Примечание
Амортизационные затраты на спецоборудование	4 769	Таблица 8
Затраты на основную заработную плату	125 913	Таблица 10
Затраты на дополнительную заработную плату	18 886,95	Таблица 11

Затраты на отчисление во внебюджетные фонды	40 543	Таблица 12
Накладные расходы	30 417,9	16% от суммы статей 1-4
Бюджет затрат НИИ	220 529,86	Сумма статей 1-5

5.3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности проекта происходит на основе расчета интегрального показателя эффективности научного исследования. Основные показатели следующие:

Интегральный показатель финансовой эффективности получают в ходе оценки бюджета затрат нескольких вариантов выполнения исследования.

Интегральный финансовый показатель определяется так:

$$I_{\text{фин } p}^{\text{исп } i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}},$$

Где $I_{\text{фин } p}^{\text{исп } i}$ – интегральный финансовый показатель разработки;

Φ_{pi} – стоимость i -го варианта исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта.

Интегральный показатель ресурсоэффективности определяется следующим образом:

$$I_{pi} = \sum a_i * b_i$$

Где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i - бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Таблица 14 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ Объект исследования	Весовой коэффициент параметра	Исп. 1	Исп. 2	Исп. 3
Способствует росту производительности труда пользователя	0,2	5	3	4
Удобство в эксплуатации (соответствует требованиям потребителей)	0,3	4	3	3
Помехоустойчивость	0,1	4	5	2
Энергосбережение	0,1	4	5	3
Надежность	0,2	3	5	5
Материалоемкость	0,1	4	4	5
Итого:	1			

$$I_{p-ucn1} = 5 * 0,2 + 4 * 0,3 + 4 * 0,1 + 4 * 0,1 + 3 * 0,2 + 4 * 0,1 = 4$$

$$I_{p-ucn2} = 3 * 0,2 + 3 * 0,3 + 5 * 0,1 + 5 * 0,1 + 5 * 0,2 + 4 * 0,1 = 3,9$$

$$I_{p-ucn3} = 4 * 0,2 + 3 * 0,3 + 2 * 0,1 + 3 * 0,1 + 5 * 0,2 + 5 * 0,1 = 3,7$$

Интегральный показатель эффективности вариантом использования определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{испi} = \frac{I_{p-испi}}{I_{финр.i}}$$

Сравнительная эффективность проекта:

$$\mathcal{E}_{cp} = \frac{I_{исп1}}{I_{исп2}}$$

Таблица 11 - Сравнительная эффективность разработки

№	Показатели	Исп. 1	Исп. 2	Исп. 3
1	Интегральный финансовый показатель разработки	1	0,95	0,9
2	Интегральный показатель ресурсоэффективности разработки	4	3,9	3,7
3	Интегральный показатель эффективности	4	4,1	4,11
4	Сравнительная эффективность вариантов исполнения	0,97	0,99	1

В данном разделе были определены финансовый показатель разработки, показатель ресурсоэффективности, интегральный показатель эффективности и, на основании сравнительной эффективности вариантов исполнения, оптимальным был выбран 3 вариант исполнения.

Глава 6. Социальная ответственность

Научно-исследовательская работа представляет собой алгоритмическое и программное обеспечение поиска и анализа данных социальных сетей для составления обобщенного портрета студента Томского Политехнического Университета. Программный продукт позволяет установить не всегда очевидные закономерности между данными в социальных сетях и некоторыми факторами. В рамках предлагаемой работы будут спроектированы и реализованы комплекс алгоритмов и программных модулей для автоматизированного параметризованного поиска, извлечения, хранения и анализа открытой информации социальных сетей. Прикладное назначение разрабатываемой системы – информационная поддержка маркетинговых, социологических и политологических исследований. Приоритетным источником данных является социальная сеть ВКонтакте.

Практически вся работа будет осуществляться на компьютере, который является источником большого количества вреда для здоровья человека. Экран компьютера представляет сразу две опасности: синий свет при длительном пользовании приводит к раздражительности и нарушению сна, а постоянная концентрация на тексте или мелких деталях может вызвать ухудшение зрения. Неподвижная поза в течение продолжительного времени способствует появлению болевых ощущений в области спины, плечевых суставов и шеи.

6.1. Правовые и организационные вопросы обеспечения безопасности

6.1.1. Организационные мероприятия обеспечения безопасности

Рабочее место должно быть организовано с учетом требований ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» и СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».

Некоторые, наиболее важные пункты из ГОСТ (нумерация пунктов сохранена в соответствии с ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования»):

2.4.1. Конструкция регулируемого кресла оператора должна соответствовать требованиям ГОСТ 21889-76.

4.2. Очень часто используемые средства отображения информации, требующие точного и быстрого считывания показаний, следует располагать в вертикальной плоскости под углом $\pm 15^\circ$ от нормальной линии взгляда и в горизонтальной плоскости под углом $\pm 15^\circ$ от сагиттальной плоскости.

СанПиН (нумерация пунктов сохранена в соответствии с СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы»):

3.2. Естественное и искусственное освещение должно соответствовать требованиям действующей нормативной документации. Окна в помещениях, где эксплуатируется вычислительная техника, преимущественно должны быть ориентированы на север и северо-восток.

3.4. Площадь на одно рабочее место пользователей ПЭВМ с ВДТ на базе электронно-лучевой трубки (ЭЛТ) должна составлять не менее 6 м СанПиН 2.2.2/2.4.1340-03, в помещениях культурно-развлекательных учреждений и с ВДТ на базе плоских дискретных экранов (жидкокристаллические, плазменные) - 4,5 м СанПиН 2.2.2/2.4.1340-03.

3.7. Помещения, где размещаются рабочие места с ПЭВМ, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации.

4.2. В производственных помещениях, в которых работа с использованием ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты

управления, залы вычислительной техники и др.) и связана с нервно-эмоциональным напряжением, должны обеспечиваться оптимальные параметры микроклимата для категории работ 1а и 1б в соответствии с действующими санитарно-эпидемиологическими нормативами микроклимата производственных помещений. На других рабочих местах следует поддерживать параметры микроклимата на допустимом уровне, соответствующем требованиям указанных выше нормативов.

5.1. В производственных помещениях при выполнении основных или вспомогательных работ с использованием ПЭВМ уровни шума на рабочих местах не должны превышать предельно допустимых значений, установленных для данных видов работ в соответствии с действующими санитарно-эпидемиологическими нормативами.

6.3. Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300-500 лк. Освещение не должно создавать бликов на поверхности экрана. Освещенность поверхности экрана не должна быть более 300 лк.

9.4. Экран видеомонитора должен находиться от глаз пользователя на расстоянии 600-700 мм, но не ближе 500 мм с учетом размеров алфавитно-цифровых знаков и символов.

10.1. Высота рабочей поверхности стола для взрослых пользователей должна регулироваться в пределах 680-800 мм; при отсутствии такой возможности высота рабочей поверхности стола должна составлять 725 мм.

10.6. Клавиатуру следует располагать на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю, или на специальной, регулируемой по высоте рабочей поверхности, отделенной от основной столешницы.

6.1.2. Особенности законодательного регулирования проектных решений

На основании правил пользования Сайтом ВКонтакте, при разработке приложения для анализа данных из социальных сетей, необходимо учесть следующие пункты (нумерация сохранена):

6.3.9. Запрещается использовать без специального на то разрешения Администрации Сайта автоматизированные скрипты (программы, боты, краулеры) для сбора информации на Сайте и/или взаимодействия с Сайтом и его функционалом;

6.3.11. Запрещается осуществлять незаконные сбор и обработку персональных данных других лиц;

6.7. Создаваемые Пользователями приложения API должны использовать только опубликованные на Сайте методы API, а также ID, защищенный ключ и сервисный ключ доступа, указанные в настройках данных приложений. Использование других методов API, а также ID, защищенного ключа и сервисного ключа доступа приложений API третьих лиц, в т.ч. приложения API Администрации Сайта, строго запрещено. Пользователь обязуется регулярно проверять перечень разрешённых методов и незамедлительно вносить корректировки в функциональность своих приложений API в соответствии с изменениями перечня. За нарушение настоящего пункта Пользователь несет предусмотренную применимым законодательством, настоящими Правилами и иными документами Администрации Сайта ответственность. Администрация Сайта при этом оставляет за собой право на защиту собственных прав и законных интересов.

В результате рассмотрения пользовательского соглашения был сделан вывод, что использование скриптов, ботов и программ без разрешения администрации «Вконтакте» запрещено, а разрешённые к использованию методы API должны быть опубликованы на сайте данной социальной сети.

6.2. Производственная безопасность

В данном пункте производится анализ вредных и опасных факторов, которые могут возникнуть на одном из этапов выполнения работы.

Стоит заметить, что химические факторы и шумы не оказывают существенное влияние на состояние здоровья человека, поэтому данные пункты не рассматриваются.

Таблица 1 – Возможные опасные и вредные факторы

Факторы (по ГОСТ 12.0.003-74)	Этапы работ			Нормативные документы
	Проектирование	Программирование	Эксплуатация	
Отклонение показателей микроклимата	+	+	+	СанПиН 2.2.2.548-96
Отсутствие или недостаток естественного света	+	+	+	СНиП 23-05-95*
Недостаточная освещенность рабочей зоны	+	+	+	СанПиН 2.2.1/2.1.1.1278-03
Повышенный уровень электромагнитных излучений	+	+	+	СанПиН 2.2.4/2.1.8.055-96.

6.2.1. Анализ опасных и вредных производственных факторов

Таблица 2 – Влияние опасных и вредных факторов

Фактор	Источник	Воздействие	Допустимые нормы
Отклонение показателей микроклимата	Кондиционеры/их отсутствие, увлажнители воздуха/их отсутствие	Вялость, усталость, сниженная концентрация и внимание	Таблица 3
Отсутствие или недостаток естественного света	Отсутствие окон	Ухудшение зрения, усталость глаз	КЕО не ниже 1,2%-1,5%
Недостаточная освещенность рабочей зоны	Отсутствие или неподходящие лампы	Ухудшение зрения, усталость глаз	Освещенность на рабочей поверхности от системы общего искусственного освещения 200-300 лк.
Повышенный уровень электромагнитных излучений	Все части компьютера	Возможно возникновение рака	Напряженность электростатического поля не более 20 кВ/м

Для описания допустимых норм микроклимата невозможно ограничиться одним параметром, так как микроклимат включает в себя температуру помещения, температуру поверхностей, влажность воздуха и скорость движения воздуха.

Таблица 3 – Допустимые величины показателей микроклимата по СанПиН 2.2.2.548-96

Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	20,0-21,9	19,0-26,0	15-75	0,1
Теплый	21,0-22,9	20,0-29,0		

6.2.2. Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя (работающего)

Для восстановления и поддержания допустимого микроклимата необходимо придерживаться следующих правил:

- Оборудование помещения системами обогрева, вентилирования и увлажнения
- Защита фасада здания от солнца: шторы, жалюзи, навесы и т.д.
- Рационально размещать рабочие места
- Своевременная влажная уборка помещения

Для решения проблемы отсутствия или недостатка естественного света и плохой освещенности рабочего места подходят следующие пункты:

- Сокращение времени работы
- Своевременная чистка стекол в светопроемах
- Снос деревьев, препятствующих проникновению света в помещение
- Ремонт помещения в светлых тонах
- Установка более мощных ламп или в большем количестве
- Установка ламп в правильном положении

Повышенный уровень электромагнитных излучений можно избежать, если следовать следующим пунктам:

- выключать монитор каждый раз, когда отходите от рабочего места;

- использовать монитор с жидкокристаллическим экраном;
- располагать монитор в углу помещения для того, чтобы стены поглощали излучение;
- выключать компьютер при его неиспользовании;
- сокращать время, проводимое за компьютером.

6.3. Экологическая безопасность

На данный момент персональные компьютеры не оказывают сильного воздействия на окружающую среду, однако, многое сырье, необходимое для сборки машины, является токсичным.

Именно неправильная утилизация таких отходов, как макулатура и канцелярия, детали персонального компьютера и печатающих устройств, батарейки и люминесцентные лампы ведут к загрязнению гидросферы и литосферы.

Бумагу и канцелярские принадлежности следует сортировать и утилизировать надлежащим образом. Переработанные отходы могут послужить вторично, что экономит природные ресурсы, то есть сохранит леса.

Особое внимание стоит уделить люминесцентным лампам, которые содержат в себе ртуть. Их, а также батарейки необходимо сдавать компаниям для переработки подобных материалов.

В настоящее время экология нашей планеты находится в критическом состоянии, поэтому предприятиям и всем людям необходимо придерживаться следующих правил для уменьшения негативного влияния на природу:

- Использовать энергосберегающие лампы
- Экономно использовать электроэнергию
- Экономить воду: не оставлять воду включенной при чистке зубов/бритье, содержать сантехнику в порядке
- Сдавать бумагу и пластиковые бутылки в пункты переработки

- Утилизировать батарейки, люминесцентные лампы и технику надлежащим образом

6.3.1. Безопасность в чрезвычайных ситуациях

Чрезвычайная ситуация (ЧС) – обстановка на определенной территории, сложившаяся в результате аварии, опасного природного явления, катастрофы, стихийного или иного бедствия, которая может повлечь за собой человеческие жертвы, а также ущерб здоровью человека или окружающей среде, значительные материальные потери и нарушение условий жизнедеятельности.

В помещении, где выполняется работа, наиболее вероятно возникновение пожара. Возможные причины: перегрузка в электросети, короткое замыкание, разрушение изоляции проводников, а также из-за несоблюдения техники пожарной безопасности.

Во избежание подобной ситуации необходимо проводить следующие пожарно-профилактические мероприятия:

- Периодические инструктажи работающему персоналу
- Издание инструктажей, плакатов и планов эвакуации
- Соблюдение противопожарных норм и правил при монтаже электропроводок, установке техники, систем поддержания микроклимата
- Обязательное наличие систем противопожарной сигнализации и огнетушителей

Действия в случае возникновения пожара:

1. Определите для себя, выходить или не выходить наружу. Убедитесь, что за дверью нет пожара, приложив свою руку к двери или к металлической ручке. Если они горячие, то ни в коем случае не открывайте дверь.
7. Не входите туда, где большая концентрация дыма и видимость менее 10 метров.

Если дым и пламя позволяют выйти из помещения (здания) наружу:

1. Уходите скорее от огня, используя основные и запасные пути эвакуации.
8. Отключите попутно электроэнергию.
9. Идите к выходу на четвереньках, так как вредные продукты горения скапливаются на уровне нашего роста и выше, закрывая при этом рот и нос подручными средствами защиты.
10. По пути за собой плотно закрывайте дверь.
11. Покинув опасное помещение, не вздумайте возвращаться назад, сообщите о себе должностным лицам.

Если дым и пламя в соседних помещениях не позволяет выйти наружу:

1. Не поддавайтесь панике.
12. Накройтесь полностью мокрым покрывалом (тканью).
13. Проверьте существует ли возможность выйти на крышу или спуститься по пожарной лестнице.
14. Если возможности эвакуироваться нет, то для защиты от тепла и дыма необходимо надёжно загерметизировать своё помещение:
 - а) плотно закройте входную дверь, заткните щели двери изнутри помещения, используя при этом любую ткань;
 - б) закройте окна, форточки, заткните вентиляционные отверстия;
 - с) если есть вода, постоянно смачивайте дверь, пол.
15. Если помещение наполнилось дымом, передвигайтесь на четвереньках, прикрыв рот и нос влажной тряпкой (носовым платком, рукавом от рубашки), в сторону окна и находитесь возле окна и привлекайте к себе внимание людей на улице.

Если у вас телефон, то обязательно позвоните «112», «01» (с сотового тел. 01*, 112) и сообщите, где вы находитесь. Ни в коем случае не открывайте и не разбивайте окна, так как нарушится герметичность вашего помещения, что приведёт к увеличению температуры и площади пожара.

Вывод по разделу «Социальная ответственность»

В заключении данного раздела, можно сделать вывод, что рабочее помещение, в котором выполняется работа над проектом, удовлетворяет всем необходимым нормам, что способствует продуктивному и безопасному рабочему процессу. Особое внимание хочется обратить на соответствие проведения исследования всем правовым нормам.

Заключение

В данной работе был проведен анализ студентов Томского Политехнического Университета, зарегистрированных в социальной сети ВКонтакте и вступивших в официальное сообщество университета.

Собранная информация была обработана, и лишние данные (удаленные или заблокированные пользователи) удалены из массива данных. Таким образом, для анализа были предоставлены данные 18426 пользователей.

С помощью разработанной программы были построены следующие круговые диаграммы: соотношение мужского и женского пола, география проживания студентов (страна, город), семейное положение, наличие информации о книгах, цитатах и интересах.

На основании всех исследований можно сделать вывод, что большинство студентов Томского Политехнического Университета стараются не указывать информацию о себе в социальных сетях, поэтому стоит рассматривать неструктурированный контент пользователя, такой как записи на стене и комментарии, связанные с Томским Политехническим Университетом.

Список использованных источников

1. Statista [Электронный ресурс]. Режим доступа: <https://www.statista.com> (дата обращения 20.03.2019).
2. We are social [Электронный ресурс] Digital 2019: global internet use accelerates. Режим доступа: <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates> (дата обращения 20.03.2019).
3. Wersm [Электронный ресурс] How much data is generated every minute on social media Режим доступа: <https://wersm.com/how-much-data-is-generated-every-minute-on-social-media/> (дата обращения 25.03.2019)
4. ХХ2век [Электронный ресурс] Сколько информации мир генерирует каждую минуту? Режим доступа: <https://22century.ru/popular-science-publications/data-never-sleeps> (дата обращения 2.04.2019)
5. Марко Бонцанини Анализ социальных медиа на Python – Перевод с английского языка. – изд. ДМК, 2018 – 288 с.
6. Томский Политехнический Университет [Электронный ресурс] JOIN.TPU: более 20 тысяч абитуриентов из России и стран СНГ играют в студентов Томского политеха. Режим доступа: <https://news.tpu.ru/news/2016/08/02/25640/> (дата обращения 16.05.2019)
7. VC.RU [Электронный ресурс] Исследование: 55% рекрутеров могут изменить решение о приеме на работу на основе профиля в соцсетях. Режим доступа: <https://vc.ru/flood/5760-social-recruiting> (дата обращения 16.05.2019)

Приложение А. Листинг программы

```
#!/usr/bin/env python3
# vim: set ai et ts=4 sw=4:

import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import datetime as dt
import csv

data_names = ['Мужской', 'Женский']
data_values = [10035, 8374]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Соотношение мужского и женского пола')

xs = range(len(data_names))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names) - 1)] )
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names )
fig.savefig('sex.png')

data_names_city = ['Томск', 'Кемерово', 'Москва', 'Северск', 'Другое']
data_values = [8419, 216, 556, 278, 8957]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Соотношение городов проживания студентов')

xs = range(len(data_names_city))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names_city) - 1)] )
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_city )
fig.savefig('city.png')

data_names_country =
['Россия', 'Другое', 'Казахстан', 'Китай', 'США', 'Япония', 'Объединенные Арабские
Эмираты']
data_values = [14439, 2896, 893, 92, 85, 16, 5]

dpi = 70
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Соотношение стран проживания студентов')
```

```

xs = range(len(data_names_country))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names_country) - 1)] )
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_country )
fig.savefig('country.png')

data_names_rel = ['В браке/в отношениях', 'Свободны', 'Не указано']
data_values = [1262, 1057, 15662]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Соотношение студентов по семейному положению')

xs = range(len(data_names_rel))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names_rel) - 1)] )
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_rel )
fig.savefig('rel.png')

data_names_book = ['Не указаны', 'Указаны']
data_values = [17750, 676]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Любимые книги')

xs = range(len(data_names_book))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names_book) - 1)])
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_book )
fig.savefig('book.png')

data_names_interes = ['Не указаны', 'Указаны']
data_values = [17488, 938]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Интересы')

xs = range(len(data_names_interes))

plt.pie(
    data_values, autopct='%.1f', radius = 1.1,

```



```

    explode = [0.15] + [0 for _ in range(len(data_names_interes) - 1)]
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_interes )
fig.savefig('interes.png')

data_names_phrases= ['Нет', 'Есть']
data_values = [17562, 863]

dpi = 80
fig = plt.figure(dpi = dpi, figsize = (512 / dpi, 384 / dpi) )
mpl.rcParams.update({'font.size': 9})

plt.title('Цитаты')

xs = range(len(data_names_phrases))

plt.pie(
    data_values, autopct='%1f', radius = 1.1,
    explode = [0.15] + [0 for _ in range(len(data_names_phrases) - 1)])
plt.legend(
    bbox_to_anchor = (-0.16, 0.45, 0.25, 0.25),
    loc = 'lower left', labels = data_names_phrases )
fig.savefig('phrases.png')

```