



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: 09.03.04 «Программная инженерия»

ООП/ОПОП: Разработка программно-информационных систем

Отделение школы (НОЦ): ОИТ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Тема работы
Разработка программного модуля психологической типологизации на основе обработки естественного языка

УДК 004.422.8:159.923

Обучающийся

Группа	ФИО	Подпись	Дата
8К93	Третьяков Дмитрий Андреевич		

Руководитель ВКР

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
Профессор ОСГН	Гасанов Магеррам Али оглы	Д.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ООД ШБИП	Мезенцева Ирина Леонидовна			

ДОПУСТИТЬ К ЗАЩИТЕ:

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

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ООП

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.
УК(У)-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений.
УК(У)-3	Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде.
УК(У)-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(-ых) языке(-ах).
УК(У)-5	Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах.
УК(У)-6	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни.
Общепрофессиональные компетенции	
ОПК(У)-1	Способен применять естественнонаучные и общепрофессиональные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности.
ОПК(У)-2	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности.
ОПК(У)-3	Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.
ОПК(У)-4	Способен участвовать в разработке стандартов, норм и правил, а также технической документации, связанной с профессиональной деятельностью.
ОПК(У)-5	Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем.
ОПК(У)-7	Способен осуществлять выбор платформ и инструментальных программно-аппаратных средств для реализации информационных систем.
Профессиональные компетенции	
ПК(У)-1	Способен выполнять интеграцию программных модулей и компонент.

ПК(У)-2	Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения.
ПК(У)-9	Способность использовать когнитивный подход и воспринимать (обобщать) научно-техническую информацию, отечественный и зарубежный опыт по тематике исследования.
ПК(У)-13	Способность использовать информационные технологии и инструментальные средства при разработке проектов.



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 09.03.04 Программная инженерия
Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП/ОПОП
_____ Чердынцев Е.С.
(Подпись) (Дата) (ФИО)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

Обучающийся:

Группа	ФИО
8К93	Третьяков Дмитрий Андреевич

Тема работы:

Разработка программного модуля психологической типологизации на основе обработки естественного языка	
Утверждена приказом директора (дата, номер)	№102-30/с от 12.04.2023

Срок сдачи обучающимся выполненной работы:	09.06.2023
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе (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 функционированию (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.)</p>	<p>Объектом исследования является программный модуль психологической типологизации на основе обработки естественного языка</p>
---	--

<p>Перечень разделов пояснительной записки подлежащих исследованию, проектированию и разработке (аналитический обзор литературных источников 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)</p>	<ol style="list-style-type: none"> 1. Обзор методов и подходов анализа естественного языка по методике MBTI 2. Подготовка данных, выбор и экспериментальная апробация модели психологической типологизации. Программная реализация модуля 3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение 4. Социальная ответственность
<p>Перечень графического материала (с точным указанием обязательных чертежей)</p>	<ol style="list-style-type: none"> 1. Слайды в формате pptx, демонстрирующие результаты работы
<p>Консультанты по разделам выпускной квалификационной работы (с указанием разделов)</p>	
<p>Раздел</p>	<p>Консультант</p>
<p>Финансовый менеджмент, ресурсоэффективность, ресурсосбережение</p>	<p>Профессор ОСГН, Гасанов Магеррам Али оглы</p>
<p>Социальная ответственность</p>	<p>Старший преподаватель, Мезенцева Ирина Леонидовна</p>

<p>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</p>	<p>01.03.2023</p>
--	-------------------

Задание выдал руководитель:

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

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
<p>8K93</p>	<p>Третьяков Дмитрий Андреевич</p>		



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 – 09.03.04 «Программная инженерия»
Уровень образования – Бакалавриат
Отделение школы (НОЦ) Отделение информационных технологий
Период выполнения (осенний / весенний семестр 2022 /2023 учебного года)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы**

Обучающийся:

Группа	ФИО
8К93	Третьяков Дмитрий Андреевич

Тема работы:

Разработка программного модуля психологической типологизации на основе обработки естественного языка		
Срок сдачи обучающимся выполненной работы:		09.06.2023
Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
30.03.2023	Подготовка обзора методов и подходов анализа естественного языка по методике МВТИ	20
25.04.2023	Подготовка данных, выбор и экспериментальная апробация модели психологической типологизации.	20
25.05.2023	Программная реализация модуля	30
15.05.2023	Финансовый менеджмент, ресурсоэффективность, ресурсосбережение	10
15.05.2023	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/ОПОП

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

Обучающийся

Группа	ФИО	Подпись	Дата
8К93	Третьяков Дмитрий Андреевич		

РЕФЕРАТ

Выпускная квалификационная работа 97 с., 15 рис. 28 табл., 37 источников, 3 прил.

Ключевые слова: нейросети, большие данные, глубокое обучение, анализ личности, психологическая типологизация, обработка естественного языка, тестирование, MBTI, Big Five.

Объектом исследования является модуль психологической типологизации на основе обработки естественного языка.

Цель работы является создание модуля психологической типологизации на основе обработки естественного языка.

В процессе исследования был осуществлен обзор методов и подходов анализа естественного языка. Были подготовлены данные, осуществлен выбор и последующая экспериментальная апробация моделей психологической типологизации, определена архитектура модуля, разработано веб-приложение, реализующее работу модуля.

В результате исследования был разработан модуль, обучены нейросети, проведена оценка результатов работы модуля.

Основные конструктивные, технологические и технико-эксплуатационные характеристики – использование технологий обработки естественного языка и машинного обучения.

Степень внедрения – выполнена программная реализация модуля.

Область применения: подразделения кадровых служб, рекрутинг (найм) персонала, медицина, образовательные организации.

Экономическая эффективность/значимость работы – автоматизация процесса тестирования, процедуры оценки на соответствие заданным критериям.

В будущем планируется: расширение функциональности и улучшение точности модуля, с помощью использования систем машинного обучения и искусственного интеллекта. Также планируется интеграция модуля в новые приложения и системы.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	12
ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.....	14
1. ОСНОВНОЙ РАЗДЕЛ.....	15
1.1 Анализ основных подходов, используемых для определения типов личности.....	15
1.1.1 Описание типологии MBTI.....	15
1.1.2 Описание модели Big Five («Большая пятерка»).....	16
1.2 Анализ существующих подходов к автоматизированному определению типов личности.....	20
1.2.1 Анализ больших данных.....	20
1.2.2 Использование носимых технологий.....	21
1.2.3 Игровизация.....	21
1.2.4 Видео-резюме.....	22
1.2.5 Автоматизированное тестирование личности.....	23
1.3 Подготовка среды разработки.....	24
1.4 Проектирование модели.....	27
1.4.1 Обоснование выбора модели.....	28
1.5 Входные и выходные данные.....	29
1.5.1 Датасеты MBTI.....	30
1.5.2 Датасеты Big Five.....	31
1.6 Предобработка и подготовка входных данных.....	32
1.7 Обучение моделей и оценка качества.....	34
1.8 Выводы на основе тестирования.....	37
1.9 Проектирование и реализация веб-приложения.....	39
1.9.1 Анализ требований к функциональности приложения.....	39

1.9.2	Проектирование интерфейса приложения.....	41
1.9.2.1	Проектирование архитектуры приложения.....	42
1.9.2.2	Итоговое скриншоты приложения	45
1.10	Выводы по разделу.....	48
ЗАДАНИЕ К РАЗДЕЛУ «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»		49
2.	ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	50
2.1	Потенциальные потребители результатов исследования.....	50
2.2	Анализ конкурентных технических решений	51
2.3	Технология Quad	53
2.4	SWOT-анализ	54
2.5	Планирование научно-исследовательских работ.....	57
2.5.1	Структура работ в рамках научного исследования	57
2.5.2	Определение трудоемкости выполнения работ	59
2.5.3	Разработка графика проведения научного исследования	60
2.6	Бюджет научно-исследовательских работ	63
2.6.1	Расчет материальных затрат НИР	64
2.6.2	Расчет затрат на специальное оборудование для научных (экспериментальных) целей	65
2.6.3	Основная заработная плата исполнителя темы.....	66
2.6.4	Расчет дополнительной заработной платы исполнителей темы ...	68
2.6.5	Отчисления во внебюджетные фонды (страховые отчисления) ...	69
2.6.6	Накладные расходы.....	70
2.6.7	Формирование бюджета затрат научно-исследовательского проекта 71	
2.7	Определение ресурсной (ресурсосберегающей), финансовой,	

бюджетной, социальной и экономической эффективности исследования	72
2.8 Выводы по разделу	74
ЗАДАНИЕ К РАЗДЕЛУ «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»	76
3. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	78
3.1 Введение	78
3.2 Правовые и организационные вопросы обеспечения безопасности	78
3.2.1 Правовые нормы трудового законодательства РФ	78
3.2.2 Эргономические требования к правильному расположению и компоновке рабочей зоны	79
3.3 Производственная безопасность.....	81
3.3.1 Статические физические перегрузки	81
3.3.2 Монотонный режим работы	82
3.3.3 Отсутствие или недостаток освещенности рабочей зоны	82
3.3.4 Отклонение показателей микроклимата	83
3.3.5 Опасность поражения электрическим током	84
3.4 Экологическая безопасность	85
3.4.1 Воздействие на селитебную зону	85
3.4.2 Воздействие на литосферу	86
3.4.3 Воздействие на гидросферу	86
3.4.4 Воздействие на атмосферу	86
3.5 Безопасность в чрезвычайных ситуациях	87
3.6 Выводы по разделу	88
ЗАКЛЮЧЕНИЕ	89
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	90
ПРИЛОЖЕНИЕ А Предобработка данных	95

ПРИЛОЖЕНИЕ Б Сборка модели.....	96
ПРИЛОЖЕНИЕ В Пользовательский callback на keras	97

ВВЕДЕНИЕ

В современном информационном обществе социальные сети играют огромную роль в нашей повседневной жизни. Они стали неотъемлемой частью коммуникации, информационного обмена и самовыражения. Благодаря этому широкому распространению и популярности, социальные сети стали бесценным источником данных для исследования человеческой психологии.

Объект исследования: психологическая типологизация на основе обработки естественного языка.

Предмет исследования: использование данных из социальных сетей для определения личностных типов с помощью методов обработки естественного языка.

Социальные сети содержат огромное количество текстовых данных, которые отражают мысли, эмоции, интересы и взгляды пользователей. Эти данные представляют собой ценную информацию для понимания личностных особенностей и психологических состояний людей. Психологическая типологизация, основанная на обработке естественного языка, становится все более актуальной и востребованной в настоящее время.

Определение личностных типов имеет огромное значение в различных сферах жизни, таких как подбор персонала, развитие лидерских качеств, управление командой, психологическая поддержка и многое другое. Автоматизированный метод определения личностных типов с использованием данных из социальных сетей позволяет получить непосредственный и достоверный доступ к информации о личности, не требуя больших затрат времени и усилий.

Более того, психологическая типологизация на основе обработки естественного языка позволяет преодолеть некоторые ограничения традиционных методик, таких как субъективность в интерпретации результатов, сложность процесса определения типа и ограниченность выборки. Анализ текстовых данных позволяет проводить объективный и

автоматизированный анализ, основанный на статистических и алгоритмических методах.

Таким образом, психологическая типологизация на основе обработки естественного языка является актуальным и перспективным направлением исследований, которое не только упрощает процесс определения личностных типов, но и предоставляет новые возможности для более глубокого понимания человеческой психологии и развития эффективных практических подходов в различных областях деятельности.

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ИС – информационная система.

ИТ – информационные технологии.

ОС – операционная система.

ПО – программное обеспечение.

ООП – объектно-ориентированное программирование.

Датасет - то механизм хранения информации, который предоставляет быстрый доступ к большим объемам данных

MBTI - «Myers-Briggs type indicator» (индикатор типов Майерс-Бригс).

GPU - Графический процессор (GPU или Graphics Processing Unit)

OS – Operating System – операционная система.

API – Application Programming Interface – программный интерфейс приложения, интерфейс прикладного программирования.

SDK – Software Development Kit – набор средств разработки.

HTTP – HyperText Transport Protocol – протокол передачи гипертекста.

URI – Universal Resource Identifier – универсальный идентификатор ресурса.

Big Five - модель личности по пяти шкалам: открытость, добросовестность, экстраверсия, дружелюбие и нейротизм.

MAE - Mean absolute error – среднее значение абсолютной разницы между прогнозами и фактическими наблюдениями

MSE – Mean squared error – среднеквадратичная ошибка, мера разницы между прогнозами и фактическими наблюдениями.

RMSE - Root mean squared error - квадратный корень из среднеквадратичной ошибки, мера относительной точности прогнозов.

MAPE - Mean absolute percentage error- средняя абсолютная процентная ошибка, мера относительной точности прогнозов.

DEBERTa- Deep Bidirectional Encoder Representations from Transformers – глубокая двунаправленная модель кодирования представлений на основе трансформеров, разработанная для обработки естественного языка.

1. ОСНОВНОЙ РАЗДЕЛ

1.1 Анализ основных подходов, используемых для определения типов личности

1.1.1 Описание типологии MBTI

Myers-Briggs Type Indicator (MBTI) представляет собой инструмент, используемый для классификации личностных типов [1][5]. Он основывается на четырех дихотомиях, отражающих предпочтительные способы взаимодействия и принятия решений у индивидов. В современном понимании типологии MBTI, следующие дихотомии определяют личностные предпочтения:

- Экстраверсия (Extraversion, E) против Интроверсии (Introversion, I): эта дихотомия отражает предпочтительный источник энергии и фокус внимания. Экстраверты обращают свое внимание на внешний мир и получают энергию от общения с людьми и внешней стимуляции. Интроверты, напротив, предпочитают обращать свое внимание внутрь, вглядываться в свои мысли и чувства, и находят восстановление энергии в одиночестве и внутренних размышлениях.

- Ощущение (Sensing, S) против Интуиции (Intuition, N): эта дихотомия отражает предпочтительный способ восприятия информации. Люди, предпочитающие ощущение, обращают внимание на конкретные факты, детали и реальные события, руководствуясь информацией, полученной через пять органов чувств. Люди, предпочитающие интуицию, ориентируются на восприятие широких паттернов, скрытых значений и возможностей, улавливая связи и предсказывая будущие события.

- Мышление (Thinking, T) против Чувства (Feeling, F): эта дихотомия отражает предпочтительный способ принятия решений. Люди, отдающие предпочтение мышлению, стремятся к объективному анализу, логике и последовательности в принятии решений. Они основывают свои

решения на объективных критериях и правилах. Люди, отдающие предпочтение чувству, ориентируются на эмоциональные аспекты и личные ценности в принятии решений, уделяя внимание влиянию на эмоциональное состояние окружающих.

- Суждение (Judging, J) против Восприятия (Perceiving, P): эта дихотомия отражает предпочтительный способ организации жизни и принятия решений. Люди, предпочитающие суждение, склонны к структурированию своей жизни, организации и планированию. Они стремятся к определенности, установленным срокам и завершению задач. Люди, предпочитающие восприятие, более гибки и открыты для новых информации и возможностей. Они предпочитают адаптироваться к ситуациям, сохраняя открытость для новых вариантов и впечатлений.

Методика MBTI предоставляет систему классификации личностных типов, основанную на этих четырех критериях. Комбинации предпочтений по каждому критерию формируют 16 возможных типов личности в рамках MBTI, каждый из которых имеет свои уникальные особенности и характеристики.

1.1.2 Описание модели Big Five («Большая пятерка»)

В настоящее время модель Big Five является одной из наиболее авторитетных и подтвержденных эмпирически теорий личности [3]. Она основана на факторном анализе, статистическом методе, позволяющем выявлять основные измерения или факторы, лежащие в основе множества переменных. Используя этот метод, исследователи смогли сократить список тысяч психологических характеристик до пяти общих факторов личности, которые объясняют значительную часть индивидуальных различий в поведении и самосознании.

Модель «Большой пятерки» предполагает, что личность человека может быть описана с помощью пяти широких черт: экстраверсии, доброжелательности, нейротизма, открытости к опыту и добросовестности.

Каждая из этих черт представляет собой биполярный континуум, на котором человек может занимать различное положение в зависимости от степени выраженности данной черты. Таким образом, модель Big Five не дает категориальных типов личности, а предоставляет непрерывную шкалу для оценки индивидуальных различий.

Расшифруем и детализируем черты личности по модели «Большой пятерки»:

Экстраверсия — это степень общительности, разговорчивости, энергичности и социальной активности человека. Экстраверты предпочитают взаимодействовать с другими людьми, получать стимулы из внешнего мира и выражать свои эмоции открыто. Экстраверсия влияет на поведение и жизненные результаты человека, такие как профессиональный выбор, успех, здоровье и счастье.

- Высокая экстраверсия: Люди с высокой экстраверсией обычно очень общительны, дружелюбны, жизнерадостны и любят быть в центре внимания.
- Средняя экстраверсия: Люди со средней экстраверсией обычно достаточно общительны, но не излишне. Они могут наслаждаться как компанией других людей, так и уединением.
- Низкая экстраверсия: Люди с низкой экстраверсией обычно очень сдержанные, замкнутые, спокойные и предпочитают уединение. Они не любят заводить новые знакомства, адаптироваться к разным ситуациям и находить общий язык с разными людьми

Доброжелательность — это степень дружелюбия, сотрудничества, доверия и альтруизма человека. Доброжелательные люди склонны быть толерантными, сопереживающими, щедрыми и готовыми помогать другим. Доброжелательность влияет на поведение и жизненные результаты человека, такие как качество межличностных отношений, карьера, лидерство и благополучие.

- **Высокая доброжелательность:** Люди с высокой доброжелательностью обычно очень дружелюбны, сотрудничают, доверяют и альтруистичны.

- **Средняя доброжелательность:** Люди со средней доброжелательностью обычно достаточно дружелюбны, сотрудничают, доверяют и альтруистичны, но не излишне. Они могут находить общий язык с другими людьми, но не теряют своих интересов и принципов.

- **Низкая доброжелательность:** Люди с низкой доброжелательностью обычно очень недружелюбны, не сотрудничают, не доверяют и эгоистичны. Они трудно находят общий язык с другими людьми, не уважают их мнения и чувства, не готовы помогать и поддерживать

Нейротизм — это степень эмоциональной неустойчивости, тревожности, пессимизма и негативности человека. Нейротические люди склонны переживать стресс, страх, гнев, вину и депрессию. Нейротизм влияет на поведение и жизненные результаты человека, такие как психическое здоровье, физическое здоровье и самооценка.

- **Высокий нейротизм:** Люди с высоким нейротизмом обычно очень эмоционально неустойчивы, тревожны, пессимистичны и негативны. Они часто переживают стресс, страх, гнев, вину и депрессию.

- **Средний нейротизм:** Люди со средним нейротизмом обычно достаточно эмоционально неустойчивы, тревожны, пессимистичны и негативны, но не излишне.

- **Низкий нейротизм:** Люди с низким нейротизмом обычно очень эмоционально устойчивы, спокойны, оптимистичны и позитивны. Они легко справляются с трудностями и неожиданностями. Они не склонны к самокритике, имеют высокую самооценку и верят в свои способности. Они могут иметь хорошее психическое и физическое здоровье.

Открытость к опыту — это степень любознательности, творчества, интереса к новому и необычному у человека. Открытые к опыту люди склонны быть оригинальными, гибкими, фантазийными и интеллектуально активными.

Открытость к опыту влияет на поведение и жизненные результаты человека, такие как образование, культура, хобби и мировоззрение.

- Высокая открытость к опыту: Люди с высокой открытостью к опыту обычно очень любознательны, творческие, интересуются новым и необычным. Они легко усваивают новую информацию, пробуют новые вещи и способы действия, открыты для разных точек зрения и культур. Они часто занимаются интеллектуальными, художественными и духовными видами деятельности.

- Средняя открытость к опыту: Люди со средней открытостью к опыту обычно достаточно любознательны, творческие, интересуются новым и необычным, но не излишне.

- Низкая открытость к опыту: Люди с низкой открытостью к опыту обычно очень нелюбознательны, нетворческие, не интересуются новым и необычным. Они трудно усваивают новую информацию, избегают пробовать новые вещи и способы действия, закрыты для разных точек зрения и культур.

Добросовестность — это степень ответственности, организованности, целеустремленности и самоконтроля человека. Добросовестные люди склонны быть надежными, пунктуальными, трудолюбивыми и настойчивыми. Добросовестность влияет на поведение и жизненные результаты человека, такие как успех в работе, достижение целей, здоровый образ жизни и саморазвитие.

- Высокая добросовестность: Люди с высокой добросовестностью обычно очень ответственные, организованные, целеустремленные и самоконтролирующие. Они легко выполняют свои обязанности, соблюдают сроки, работают усердно и достигают своих целей.

- Средняя добросовестность: Люди со средней добросовестностью обычно достаточно ответственные, организованные, целеустремленные и самоконтролирующие, но не излишне.

- Низкая добросовестность: Люди с низкой добросовестностью обычно очень неответственные, неорганизованные, бесцельные и

несамоконтролирующие. Они редко занимаются планированием, анализом и самосовершенствованием. Они также склонны к беспорядку, небрежности и недисциплинированности.

1.2 Анализ существующих подходов к автоматизированному определению типов личности

В настоящее время проводится обширный анализ различных подходов к автоматизированному определению типов личности. В данном разделе представлены основные подходы и методы, используемые для анализа черт личности.

1.2.1 Анализ больших данных

Данный подход основывается на использовании массивных объемов данных из онлайн-источников, таких как социальные сети, веб-серфинг или электронная коммерция. Алгоритмы анализа данных выявляют черты личности на основе лингвистических, визуальных или поведенческих признаков [7]. Этот подход предоставляет возможность получить богатую и динамичную информацию о личности, однако сопряжен с этическими и методологическими проблемами, такими как конфиденциальность, валидность и надежность полученных данных [6].

Преимущества:

- Доступность большого количества данных из разных доменов и контекстов.
- Возможность извлекать сложные и скрытые закономерности из данных.
- Возможность адаптироваться к изменениям в данных и потребностях пользователей.

Недостатки:

- Нарушение прав на конфиденциальность и собственность данных.
- Низкая контролируемость и интерпретируемость алгоритмов.

- Необходимость учитывать смещения и шумы в данных.

1.2.2 Использование носимых технологий

Этот подход основывается на использовании устройств, таких как умные часы, фитнес-трекеры или датчики, для измерения физиологических или поведенческих сигналов, связанных с чертами личности, например, частота сердечных сокращений, электропроводность кожи или физическая активность. Этот подход предоставляет объективные и непрерывные данные о личности, однако сталкивается с техническими и практическими ограничениями, такими как продолжительность работы батареи, качество данных и принятие пользователями.

Преимущества:

- Точность и достоверность измерений физических параметров
- Возможность мониторинга личности в реальном времени
- Возможность интеграции с другими видами данных

Недостатки:

- Ограниченная доступность и стоимость устройств
- Низкая комфортность и мотивация к использованию устройств
- Необходимость учитывать влияние факторов окружения

1.2.3 Игровизация

Данный подход основан на использовании игр или игровых элементов, таких как очки, значки или обратная связь, для оценки черт личности на основе производительности, предпочтений или выбора. Игровизация может обеспечить увлекательные и интерактивные данные о личности, однако требует тщательного дизайна и валидации, чтобы убедиться, что игра измеряет то, что задумано, и не вводит искажений или помех.

Преимущества:

- Привлекательность и заинтересованность пользователей
- Возможность измерять не только явные, но и неявные черты личности
- Возможность создавать разнообразные и адаптивные игровые сценарии

Недостатки:

- Сложность и затратность разработки игр
- Риск потери экологической валидности и переносимости результатов
- Необходимость учитывать индивидуальные и культурные различия в игровом поведении

1.2.4 Видео-резюме

Данный подход основывается на использовании видеоматериалов кандидатов на работу для оценки их черт личности на основе вербальных и невербальных признаков, таких как тон голоса, мимика или жесты. Этот подход предоставляет естественные и целостные данные о личности, однако зависит от качества видео, контекста ситуации и интерпретации оценивающего лица.

Преимущества:

- Соответствие современным требованиям к рекрутингу и самопрезентации
- Возможность охватывать разные аспекты личности в одном формате
- Возможность использовать автоматизированные методы анализа видео

Недостатки:

- Необходимость обеспечения технического качества и безопасности видео
- Возможность влияния стереотипов и предубеждений на оценку личности
- Необходимость учитывать разные стили и культуры коммуникации

1.2.5 Автоматизированное тестирование личности

Данный подход использует компьютерные алгоритмы или искусственный интеллект для проведения и оценки тестов на личность, основанных на текстовых, голосовых или изображениях. Автоматизированное тестирование личности обеспечивает быструю и эффективную оценку черт личности, однако требует обеспечения точности, справедливости и прозрачности алгоритмов [8].

Преимущества:

- Экономия времени и ресурсов по сравнению с традиционными тестами
- Возможность адаптации тестов к индивидуальным потребностям и целям
- Возможность использования современных методов машинного обучения и анализа данных

Недостатки:

- Риск нарушения этических норм и прав человека
- Сложность проверки валидности и надежности тестов
- Необходимость учитывать социально-психологические факторы в процессе тестирования

Сравним перечисленные методы в таблице 1.

Таблица 1– Сравнение методов анализа

Метод	Преимущества	Недостатки
Анализ больших данных	Богатая и динамичная информация о личности	Этические и методологические проблемы
Использование носимых технологий	Объективные и непрерывные данные о личности	Технические и практические ограничения
Игровизация	Увлекательные и интерактивные данные о личности	Сложность дизайна и валидации игр
Видео-резюме	Естественные и целостные данные о личности	Зависимость от качества видео и интерпретации
Автоматизированное тестирование личности	Быстрая и эффективная оценка черт личности	Риск нарушения этических норм и прав человека

Каждый из этих подходов имеет свои преимущества и ограничения, исследования в этой области продолжаются с целью разработки более точных и надежных методов определения типов личности с использованием автоматизации и анализа данных.

1.3 Подготовка среды разработки

Для решения поставленной задачи было выбрано программное обеспечение Google, в частности библиотека для машинного обучения и создания нейросетей TensorFlow [9][11]. Она предоставляет мощные инструменты для создания и обучения нейронных сетей, включая модели для обработки естественного языка. TensorFlow обладает богатым функционалом, оптимизированными алгоритмами и возможностью масштабирования, что позволит эффективно реализовать модели, основанные на обработке естественного языка для типологизации. Из-за большей ассоциативности текущей задачи с data-science была выбрана интерактивная вычислительная среда Jupiter Notebook. Jupiter Notebook представляет интерактивную среду для выполнения кода, создания отчетов и визуализации данных. Она позволяет объединить код, текстовые описания и графики в одном документе, что упрощает демонстрацию и коммуникацию результатов исследования. В контексте психологической типологизации, Jupiter Notebook позволит создавать итеративные эксперименты, проводить анализ данных и визуализацию, а также документировать результаты исследования.

В настоящее время для создания нейросетей требуются большие вычислительные мощности. Для решения этой задачи используются мощности встроенных видеокарт. Для того чтобы TensorFlow корректно обнаруживал и работал с GPU моего ПК необходимо произвести ряд действий [15].

Nvidia Cuda [10] – это программа для распараллеливания вычислений, позволяющая разделять задачу на несколько мелких. В машинном обучении

центральный процессор выполняет вычисления на 1 экземпляре, а графический – на нескольких.

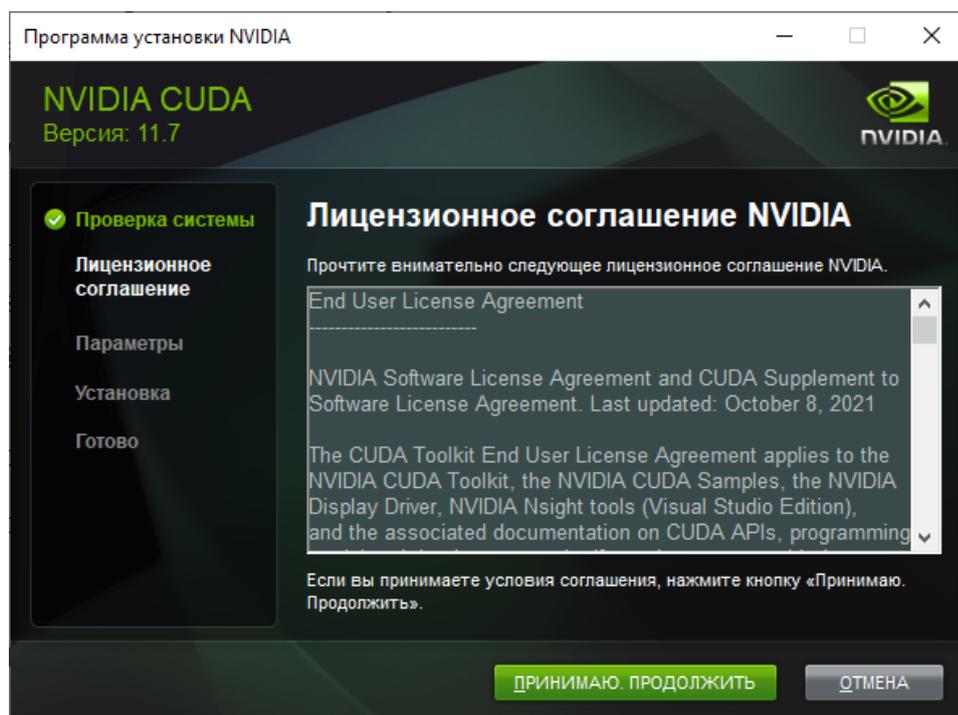


Рисунок 1 - Установка CUDA

Библиотека NVIDIA CUDA Deep Neural Network предоставляет реализации на ускорении GPU стандартных процедур в машинном обучении: свертка, нормализация и активация.

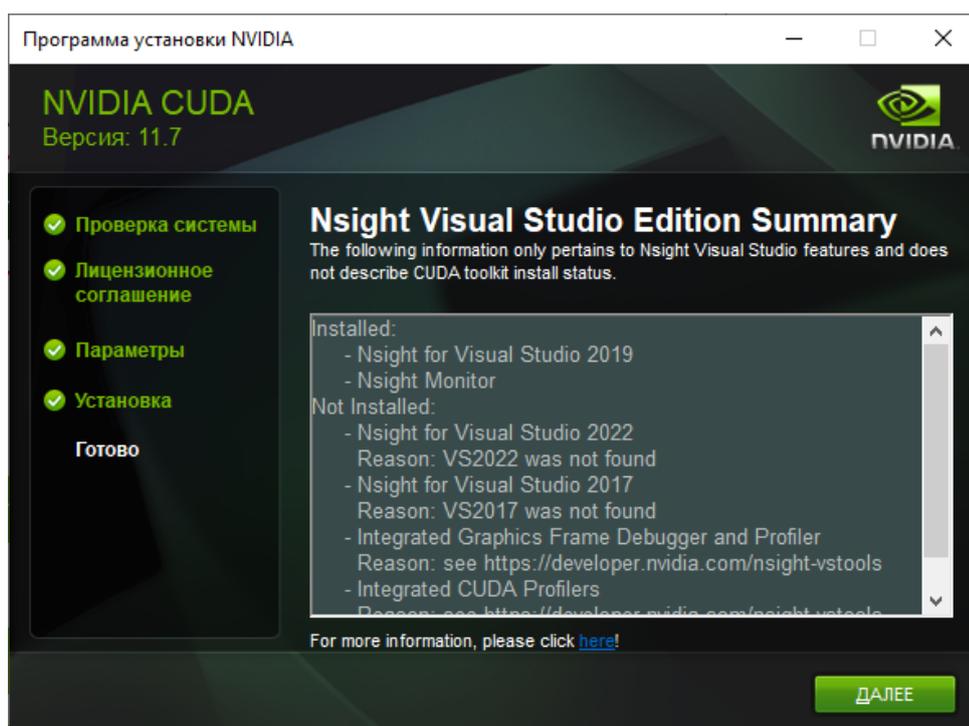


Рисунок 2 - Установка cuDNN

Visual Studio [13] – продукт разработанный Microsoft и представляет собой многофункциональную программу интегрированной среды разработки (IDE), которую можно использовать для различных аспектов разработки программного обеспечения.

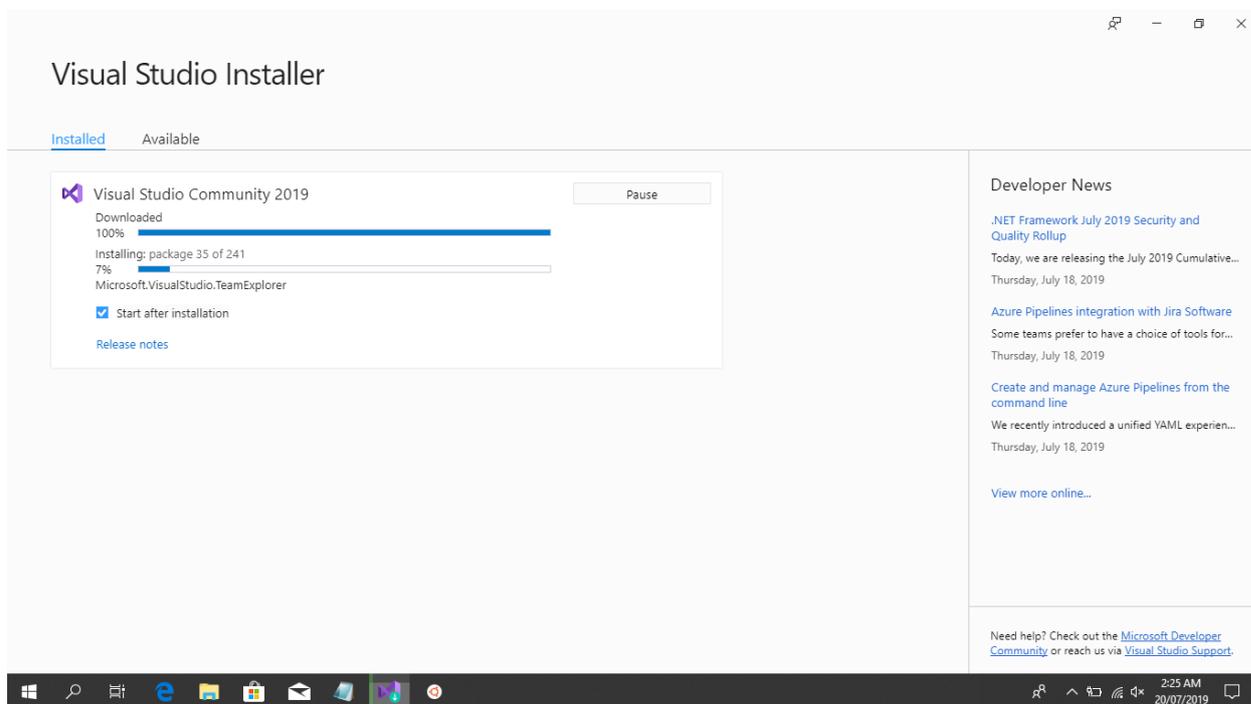


Рисунок 3 - Установка Visual Studio

Anaconda [14] – это система управления пакетами, а также менеджер среды разработки, работающий с Jupiter Notebook.

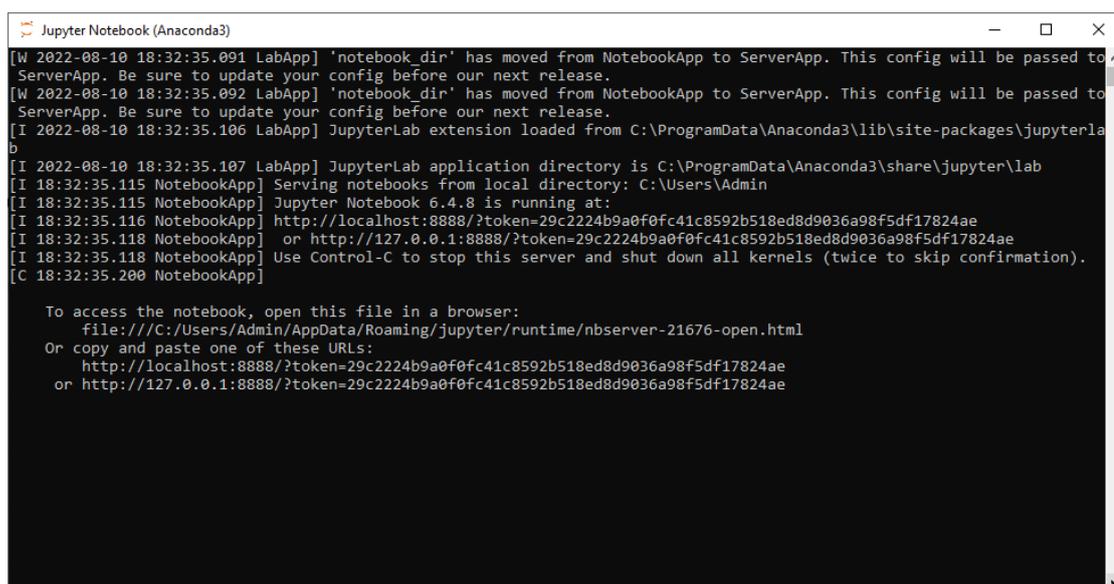


Рисунок 4 - Установка Anaconda

После установки всего необходимого, следующим действием является

проверка работоспособности программы [15].

Проверке подлежит:

- Импорт библиотеки
- Наличие GPU на устройстве (совместимость)



```
In 1 1 import tensorflow as tf
      2
      3

In 2 1 tf.test.is_gpu_available(cuda_only=False, min_cuda_compute_capability=None)

WARNING:tensorflow:From C:\Users\Admin\AppData\Local\Temp\ipykernel_9040\499444185.py:1: is_gpu_avail
depreca
ted and will be removed in a future version.
Instructions for updating:
Use `tf.config.list_physical_devices('GPU')` instead.

Out 2  True

In 2 1

In 3 1 print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
```

Num GPUs Available: 1

Рисунок 5 - Проверка работы GPU

1.4 Проектирование модели

Для решения задачи психологической типологизации на основе обработки естественного языка была выбрана архитектура нейросети на основе языковой модели DeBERTa [21]. DeBERTa — это модель глубокого обучения для задач NLP, разработанная Microsoft в 2021 году. Она использует механизмы «внимания» и выполняет задачу моделирования маскированного языка и предсказания следующего предложения, что позволяет модели лучше понимать контекст слов во входном тексте. DeBERTa также улучшает модели BERT и RoBERTa с помощью разделенного внимания, усиленного декодера маски и предварительного обучения в стиле ELECTRA с разделением вложений По градиенту.

1.4.1 Обоснование выбора модели

Модель DeBERTa [21] была выбрана среди других существующих моделей на основе нескольких критериев. Во-первых, модель DeBERTa показывает лучшие результаты на ряде стандартных бенчмарков для задач NLU и NLG по сравнению с другими моделями, такими как BERT2, RoBERTa3 или GPT-34. Например, модель DeBERTa превосходит человеческую производительность на бенчмарке SuperGLUE5 по среднему баллу (89.9 против 89.8), а также достигает лучших результатов на бенчмарках GLUE, SQuAD и RACE.

Во-вторых, модель DeBERTa имеет ряд возможностей, которые делают ее подходящей для задач мультязычного понимания текста. Одной из таких возможностей является использование межязыковых вложений (cross-lingual embeddings), которые позволяют модели кодировать тексты на разных языках в общем пространстве представлений. Это означает, что модель может сравнивать и сопоставлять тексты на разных языках без необходимости перевода. Межязыковые вложения также способствуют переносу знаний на новые языки без дополнительного обучения. Модель `m_deberta_v3` была обучена на мультязычных данных CC100, которые содержат тексты на 100 языках. Благодаря этому модель может работать с текстами на любом из этих языков или даже на смешанных языках.

В-третьих, модель DeBERTa может обрабатывать длинные текстовые последовательности, что является общей проблемой для многих задач NLP. Модель может обрабатывать до 4096 токенов за один проход, что значительно больше, чем популярные модели, такие как BERT (512 токенов) или GPT-3 (2048 токенов).

Для тонкой настройки модели была использована предобученная модель `mdeberta-v3-base`, которая имеет 12 слоев трансформеров, 768 скрытых нейронов и 12 модулей внимания. К этой модели был добавлен линейный слой с функцией активации сигмоиды для получения вероятностей

принадлежности к каждому из 16 классов по теории Майерс-Бриггс и нейрон для задачи регрессии в случае Большой пятерки. В качестве функции потерь была использована бинарная кросс-энтропия, а в качестве оптимизатора - Adam с параметром скорости обучения $5e-6$.

Архитектура модели представлена на рисунке 6:

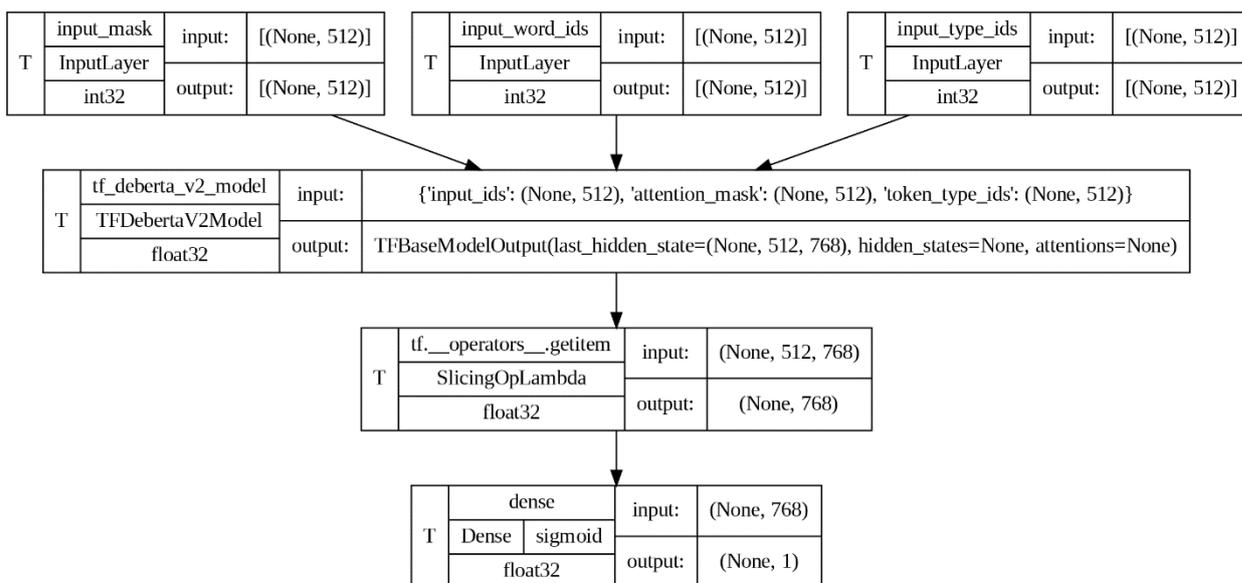


Рисунок 6 - Архитектура модели

1.5 Входные и выходные данные

Для работы были выбраны несколько датасетов, содержащих текстовые сообщения пользователей разных платформ с указанными типами личности по методике MBTI. Эти датасеты были выбраны ввиду разнообразия и большого объема данных, а также возможности сравнения результатов на разных категориях текстов. При этом данные прошли предварительную обработку: в сообщениях были заменены типы личности автора и соответствующая терминология заполнителями, что убрало тематическую предвзятость.

Датасеты, использованные в работе:

1.5.1 Датасеты MBTI

- MBTI9k - датасет, содержащий 9 тысяч сообщений пользователей с платформы Reddit [2]. Данный датасет содержит 22,934,193 комментариев и 354,996 постов. Данные разделены на четыре подмножества: посты с указанными типами личности, комментарии с указанными типами личности, подмножество комментариев авторов с более чем 1000 слов за пределами субреддитов, связанных с MBTI, и подмножество комментариев авторов с извлеченными признаками.
- MBTI Personality Type Dataset [20] - датасет, содержащий 8,675 записей постов с форума PersonalityCafe с указанными типами личности их авторов. Каждая запись содержит последние 50 постов пользователя.
- MBTI Personality Type Twitter Dataset - датасет, содержащий 7,800 записей твитов пользователей с указанными типами личности. Каждая запись содержит последние 50 твитов пользователя.
- Myers Brigg Personality Type Through Writing - датасет, содержащий 2,000 записей эссе пользователей с указанными типами личности. Каждая запись содержит одно эссе пользователя.
- Personae corpus - датасет, содержащий 145 эссе студентов (БА по лингвистике и литературе в Университете Антверпена, Бельгия) с указанными типами личности по методике MBTI. Каждое эссе имеет примерно 500 слов.
- Dutch Essays - датасет, содержащий 42 эссе студентов (БА по лингвистике и литературе в Университете Антверпена, Бельгия) с указанными типами личности по методике MBTI. Каждое эссе имеет примерно 500 слов.
- MBTI-net - датасет, содержащий 1,000 записей эссе студентов (БА по лингвистике и литературе в Университете Антверпена, Бельгия) с указанными типами личности по методике MBTI. Каждое эссе имеет примерно 500 слов.

1.5.2 Датасеты Big Five

- RusNeuroPsych [19] — это датасет, содержащий 1,500 записей результатов теста по нейропсихологии на русском языке. Каждая запись содержит текст, написанный автором в рамках теста, а также метаданные об авторе, такие как пол, возраст, оценки по психологическим тестам (Большая Пятерка, Темная Тройка, Самооценка) и предпочтения по латеральности мозга. Данные были собраны и обработаны с помощью веб-приложения RusProfiling Lab.

- MyPersonality [17] — это датасет, содержащий данные о личностных чертах, интересах и поведении миллионов пользователей Facebook. Данные были собраны с помощью популярного приложения Facebook, созданного Кембриджским центром психометрики. Пользователи могли проходить различные психометрические тесты и получать обратную связь по своим результатам. Примерно 40% респондентов также согласились поделиться данными из своего профиля Facebook с исследователями. Данные были анонимизированы и частично доступны для академических коллегаторов по всему миру через проект myPersonality. Данные содержат более 6 миллионов результатов тестов по различным шкалам (Большая Пятерка, IQ, SAT и др.) вместе с более 4 миллионами индивидуальных профилей Facebook (демография, интересы, группы, статусы и др.).

- Extroverts Tweet Differently from Introverts in Weibo [18] — это исследование, в котором с помощью машинного обучения было определено более 7,000 пользователей Weibo (китайская социальная сеть) как экстравертов или интровертов на основе их онлайн-профилей. Для этого был использован датасет Weibo Personality Corpus (WPC), который содержит 7,000 профилей пользователей Weibo с указанными типами личности по методике Большой Пятерки.

- Personality Sentence Completion Surveys Responses — это датасет, содержащий вопросы, ответы, оценки личности и метаданные, собранные из

1425 опросов по завершению предложений. Данные были размещены на OpenPsychometrics.org - некоммерческой организации, занимающейся просвещением общественности о психологии и сбором данных для психологических исследований.

1.6 Предобработка и подготовка входных данных

Входными данными для модели психологической типологизации являются текстовые сообщения пользователей разных платформ. Выходными данными являются типы личности по методике MBTI [20], которые состоят из четырехбуквенных обозначений, соответствующих двум полярным осям: интроверсия (I) - экстраверсия (E), интуиция (N) - сенсорика (S), мышление (T) - чувство (F), суждение (J) - восприятие (P). Для большой пятерки выходными данными являются оценки выраженности черты по пяти шкалам: открытость к опыту (O), сознательность (C), экстраверсия (E), доброжелательность (A) и нейротизм (N). Для каждой шкалы оценка может быть в диапазоне от 0 до 100, где 0 означает минимальную выраженность черты, а 100 - максимальную. Таким образом, модель психологической типологизации преобразует текстовые данные в числовые или категориальные метки, характеризующие личность автора текста.

В процессе обработки данных они были разделены на несколько категорий в зависимости от источника и формата текста: writing (эссе), forum (посты на форуме), reddit post (посты на Reddit), reddit comment (комментарии на Reddit), twitter (твиты). Это позволило учесть специфику каждой категории и провести сравнительный анализ результатов модели на разных категориях. Для того, чтобы подготовить данные для модели психологической типологизации, необходимо выполнить несколько шагов предобработки. Предобработка данных (фрагмент исходного кода приведен в Приложении А) позволяет улучшить качество и эффективность модели, а также уменьшить объем данных и устранить шумы и аномалии.

В данной работе были использованы следующие методы предобработки:

а) Замена URL-адресов на темы веб-страниц. URL-адреса могут содержать полезную информацию о теме или контексте сообщения. Вместо этого мы используем функцию, которая принимает на вход текст и возвращает текст с замененными URL-адресами на темы веб-страниц. Ход работы функции:

- С помощью библиотеки request отправляется запрос к веб-странице и извлекается ее заголовок (title), который обычно отражает тему страницы.

- Если заголовок не найден или запрос не выполнен, возвращается пустая строка.

- Затем заменяется URL-адрес в тексте на строку “url:” с добавлением темы страницы.

б) Обработка эмодзи. Эмодзи — это специальные символы, которые используются для выражения эмоций, настроения или отношения к сообщению. Эмодзи могут быть полезны для анализа личности, так как они отражают чувства и предпочтения автора. Однако эмодзи не являются стандартными символами естественного языка и могут быть разными на разных платформах или устройствах. Поэтому использовалась функция, которая принимает на вход текст и возвращает текст с обработанными эмодзи.

Шаги функции включали:

- Токенизацию текста на отдельные слова и символы.

- Проверку каждого токена на принадлежность к эмодзи с помощью библиотеки emoji.

- Если токен является эмодзи, мы добавляем к нему специальный тег “<EMOJI>” и его текстовое описание.

- Если токен не является эмодзи, функция добавляет его к итоговому тексту.

с) Кодирование текста для модели M-DEBERTa. DEBERTa — это модель глубокого обучения для обработки естественного языка, которая

может извлекать смысловые признаки из текста и использоваться для различных задач классификации, регрессии или генерации текста. Для того, чтобы подать текст на вход модели DEBERTa, необходимо выполнить специальное кодирование текста с помощью токенизатора DEBERTa. Токенизатор DEBERTa преобразует текст в последовательность чисел (токенов), которые соответствуют словам или частям слов из словаря DEBERTa. Функция выполняет следующие шаги:

- Токенизирует каждый текст с помощью токенизатора DEBERTa.
- Обрезает каждый текст до максимальной длины за вычетом двух специальных символов.
- Добавляет специальные символы “[CLS]” и “[SEP]” в начало и конец каждого текста.
- Конвертирует каждый токен в числовой идентификатор из словаря DEBERTa.
- Добавляет нули (падинг) до максимальной длины для каждого текста.
- Создает маску внимания для каждого текста, где единица соответствует реальному токену, а ноль - заполнителю.
- Создает тип входных данных для каждого текста, где все токены имеют значение ноль (принадлежат одному предложению).
- Возвращает словарь с ключами “input_word_ids”, “input_mask” и “input_type_ids” и значениями - массивами numpy с соответствующими параметрами.

1.7 Обучение моделей и оценка качества

Данные были разделены на обучающую (80%), валидационную (10%) и тестовую (10%) выборки. Для типологии MBTI были обучены 5 моделей: 4 модели для задач бинарной классификации (по каждому измерению) и одна модель многоклассовой классификации (по всем 16 типам). Для

классификации «по Большой пятёрки» было обучено 5 моделей регрессии по каждому измерению.

Параметры обучения моделей были следующие: размер батча - 13, количество эпох - 5, длина последовательности – 512. Для задач классификации функция потерь-бинарная перекрёстная энтропия, для классификации- средняя абсолютная ошибка. Обучение моделей производилось на графическом процессоре Tesla K80 с помощью библиотеки TensorFlow.

Для оценки качества моделей классификации на тестовой выборке были использованы метрики точности (accuracy), полноты (recall), точности (precision) и F1-меры для каждого класса. Для оценки качества моделей регрессии на тестовой выборке были использованы метрики среднеквадратичной ошибки (MSE), корневой среднеквадратичной ошибки (RMSE), средней абсолютной ошибки (MAE) и средней абсолютной процентной ошибки (MAPE) для каждой черты личности. Также для визуализации результатов классификации были построены матрицы ошибок (confusion matrix). Результаты обучения на разных наборах данных представлены в таблицах 2, 3 и 4:

Таблица 2– Результаты обучения моделей на различных задачах классификации

Model	Accuracy	Precision	Recall	F1-мера
Forum	0.8770	0.8772	0.8772	0.8772
Redditposts	0.8275	0.8267	0.8280	0.8273
Redditcomments	0.8596	0.8596	0.8595	0.8596
Writing	0.8873	0.8856	0.8900	0.8878
Twitter	0.9026	0.9030	0.9022	0.9026
Все данные	0.8700	0.8698	0.8702	0.8700

Таблица 3 – Результаты обучения для модели 16 классов

Метрика	Значение
Loss	0.1053
accuracy	0.6463
auc	0.9531
precision	0.7825
recall	0.5542
F-мера	0.6489

Таблица 4– Результаты обучения для черт Большой пятерки

Модель	MSE	RMSE	MAE	MAPE
Extraversion	0.0305	0.1747	0.0937	21.7%
Neurotizm	0.0345	0.1858	0.0997	25.9%
Openness	0.0251	0.1584	0.0741	18.9%
Concisness	0.0267	0.1635	0.0829	23.7%
Agreableness	0.0263	0.1622	0.0831	23.9%

Матрица ошибок для 16 классов приведена на рисунке 7.

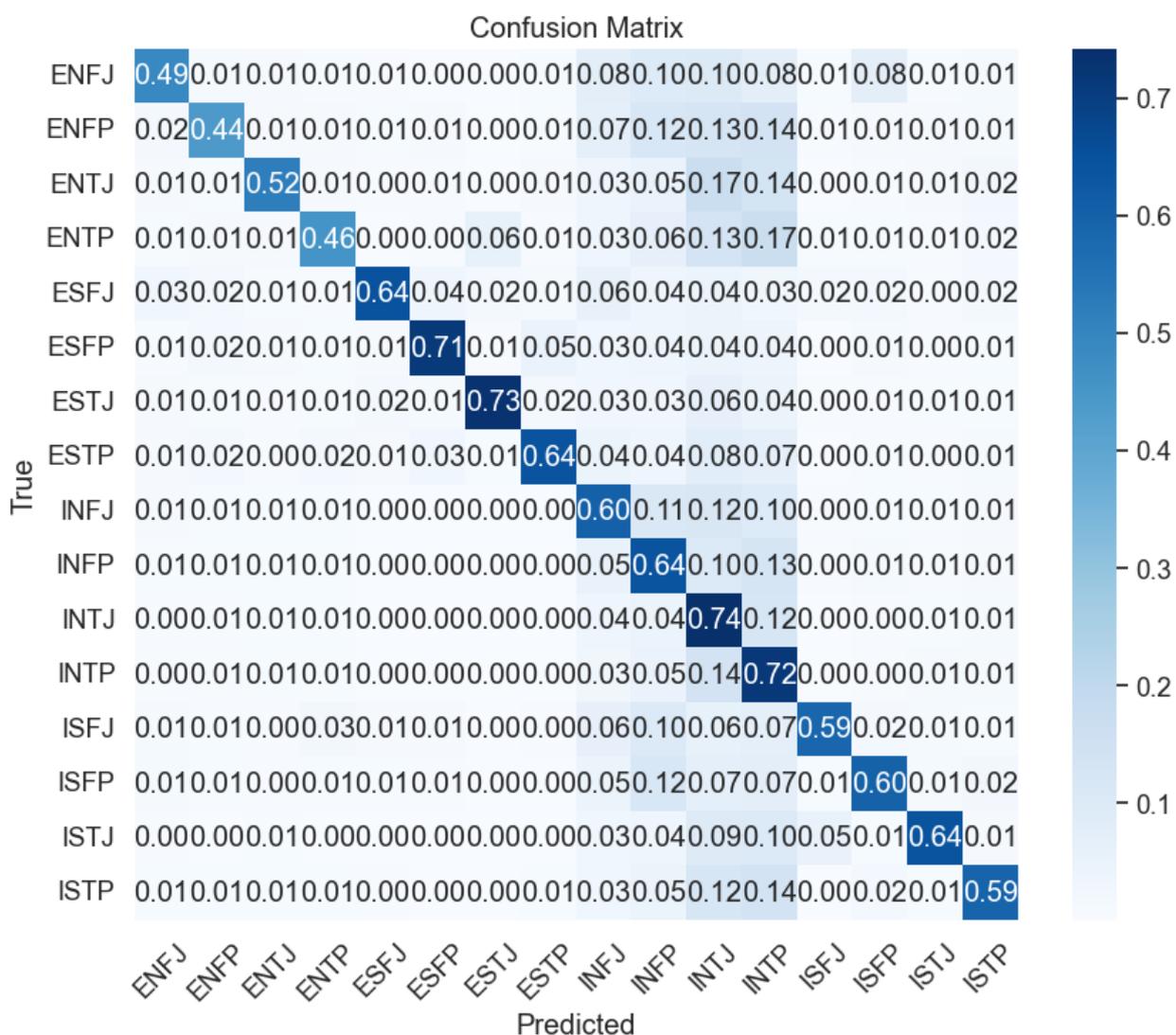


Рисунок 7 - Матрица ошибок для 16 классов

Далее представлен результат классификации одной из моделей на Экстраверсию-Интроверсию автора, с помощью алгоритма LIME [4]:

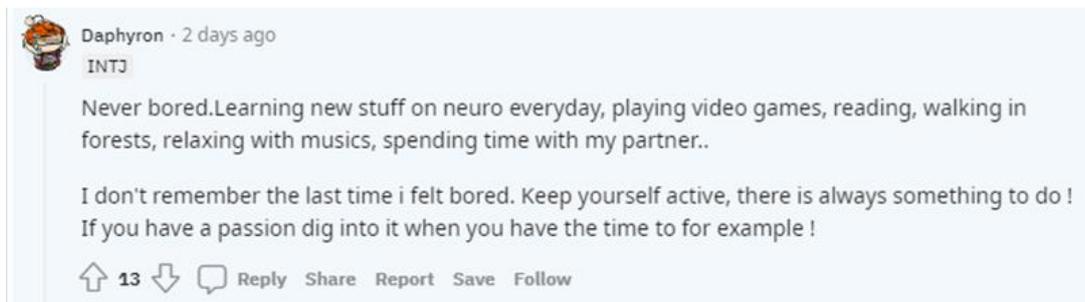


Рисунок 8 - Комментарий пользователя Reddit (Интроверт)

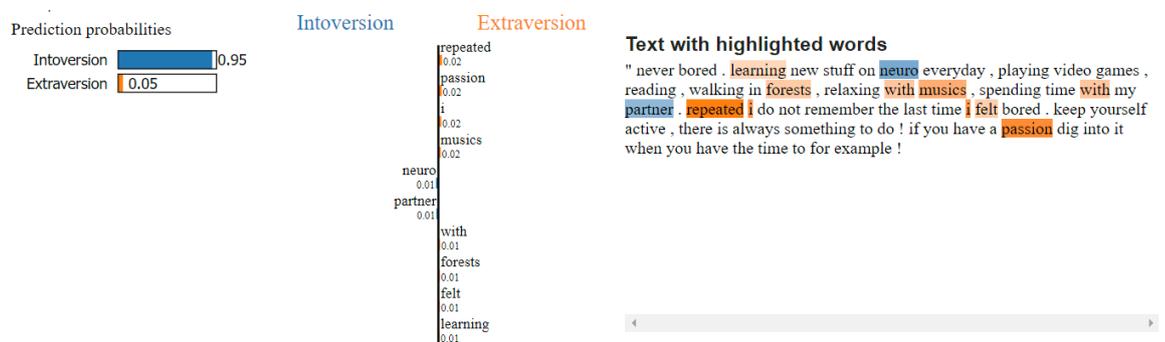


Рисунок 9 - Классификация модели и ее интерпретация с помощью LIME

В данном сообщении модель обращает внимание на уединенные занятия, что подтверждается правильностью предсказания. Фрагменты исходного кода приведены в Приложении Б и Приложении В.

1.8 Выводы на основе тестирования

В данной работе были обучены модели классификации и регрессии для определения типов и черт личности по текстам из разных источников данных. Основываясь на проведенном исследовании, можно сделать следующие выводы:

Модели классификации

- Модели классификации показали высокую точность и F1-меру на всех наборах данных, превосходящие уровень случайного угадывания (6,25%) и базовую модель, которая всегда предсказывала наиболее частый класс

(20%). Это свидетельствует о том, что модели способны правильно определять типы личности по текстам из разных источников данных.

- Для анализа типов личности мы использовали шкалу экстраверсия-интроверсия теории Юнга. Согласно Юнгу, экстраверт проявляет интерес к внешнему миру, социальным и практическим аспектам жизни, операциям с реальными объектами, а интроверт склонен к погружению в мир воображения и размышлений.

- Наилучший результат показала модель на основе данных из Twitter, которая достигла точности и F1-меры более 90% на наборе данных из твитов пользователей социальной сети Twitter. Это может объясняться тем, что твиты являются более краткими и выразительными формами текстов, которые лучше отражают индивидуальные особенности авторов.

- Наихудшая точность у модели в определении класса ENFP -44%, наилучшая точность для класса INTJ -74 %. Эти результаты показывают высокую эффективность модели, так как точность случайного угадывания для каждого класса равна в среднем 6.25%.

Модели регрессии

- Модели регрессии показали низкую ошибку при предсказании значений черт личности по текстам из социальных сетей. Ошибка моделей составила от 20 до 25 %. Это свидетельствует о том, что модели способны адекватно оценивать степень проявления черт личности по текстам из разных источников данных.

- Наилучший результат показала модель Openness, которая имеет самые низкие значения MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) и MAPE (Mean Absolute Percentage Error) среди всех моделей регрессии. Это означает, что модель Openness наиболее точно предсказывает значение черты открытости к опыту по текстам из социальных сетей.

- Остальные модели регрессии имеют схожие значения ошибок при предсказании значений черт добросовестности, экстраверсии, дружелюбия и

нейротизма. Это может говорить о том, что эти черты менее выражены или менее различимы в текстах из социальных сетей.

Таким образом, обученные модели демонстрируют соответствие классификации текстов по типу личности теории. Кроме того, можно заключить, что обученные модели демонстрируют способность оценивать степень проявления черт личности по шкале Big Five по текстам из социальных сетей с низкой ошибкой.

1.9 Проектирование и реализация веб-приложения

Для того чтобы модель применялась на реальных данных было создано веб-приложение - InnerGalaxy. Описание этапов проектирования:

На этапе проектирования веб-приложения были выполнены действия, описанные в данном разделе.

1.9.1 Анализ требований к функциональности приложения

Анализ требований к функциональности приложения. На этом этапе были определены цели, задачи и ограничения приложения. Был составлен список функциональных и нефункциональных требований к приложению, а также определены критерии качества и эффективности приложения.

Цели приложения:

- Предоставить пользователям возможность анализировать типы и черты личности по текстам из разных источников данных с помощью моделей классификации и регрессии.
- Повысить интерес и осведомленность пользователей о психологии личности и ее приложениях в разных сферах жизни.
- Способствовать развитию навыков самопознания и саморазвития у пользователей.

Ограничения приложения:

- Приложение работает только с текстами на 100 языках.
- Приложение имеет ограниченный доступ к данным из социальных сетей из-за политики конфиденциальности и безопасности.

- Приложение имеет доступ только к социальной сети ВКонтакте.

Функциональные требования:

- Приложение должно позволять пользователю вводить тексты из разных источников данных, таких как социальные сети, форумы, блоги и т.д.
- Приложение должно позволять пользователю выбирать профиль ВКонтакте для анализа по его screen_name, то есть уникальному идентификатору пользователя в социальной сети (например, id123456789 или pavel_durov).
- Приложение должно проверять, существует ли такой screen_name в ВКонтакте и является ли он публичным. Если нет, то приложение должно сообщать пользователю об ошибке и предлагать ввести другой screen_name или отказаться от анализа.
- Приложение должно извлекать тексты из профиля ВКонтакте по заданному screen_name и анализировать их на основе выбранной модели личности.
- Приложение должно показывать пользователю результаты анализа в виде графиков, таблиц и текстовых комментариев, которые объясняют, какой тип или черта личности соответствует данному тексту или автору. Например:
 - График: показывает процентное соотношение каждого фактора или дихотомии в тексте или профиле пользователя.
 - Текстовый комментарий: показывает краткое описание типа или черты личности.

Нефункциональные требования:

- Приложение не должно хранить конфиденциальные данные пользователей.

- Приложение должно поддерживать две модели анализа личности: Большая пятерка и MBTI (Майерс-Бриггс).
- Приложение должно иметь стандартные элементы интерфейса, которые обеспечивают удобство и эффективность работы пользователя с приложением.
 - Приложение должно иметь кнопки для основных действий пользователя (анализ текста, выбор профиля ВКонтакте).
 - Приложение должно иметь поля для ввода текста и screen_name пользователя.
 - Приложение должно иметь графики для показа результатов анализа по моделям Большой пятерки, которые визуализируют процентное соотношение каждого фактора или дихотомии в тексте или профиле пользователя.

Критерии качества:

- Точность моделей классификации на отдельных типах личности должна быть не менее 70% по метрике ассигасу.
- Ошибка моделей регрессии должна быть не более 25% на всех наборах данных по метрике MAPE (Mean Absolute Percentage Error).
- Время обработки одного текста одной моделью не должно превышать 10 секунд.
- Время загрузки страниц приложения не должно превышать 5 секунд, а время отклика приложения на действия пользователя не должно превышать 2 секунд.

1.9.2 Проектирование интерфейса приложения

На этом этапе были разработаны макеты страниц приложения с учетом принципов удобства использования и дизайна. Также были определены элементы управления, навигация и взаимодействие между страницами. Был выбран подходящий стиль и цветовая схема для интерфейса приложения.

1.9.2.1 Проектирование архитектуры приложения

На этом этапе была спроектирована структура приложения с учетом модульности, расширяемости и безопасности. Были определены основные компоненты приложения, такие как серверная часть, клиентская часть, база данных и внешние сервисы. Также были выбраны технологии и инструменты для реализации приложения, такие как язык программирования Python, фреймворк Flask, библиотеки для обработки естественного языка и API социальной сети ВКонтакте [16].

Серверная часть — это программа на языке Python, которая использует фреймворк Flask для обработки HTTP-запросов от клиентов и возврата HTML-страниц с результатами анализа текста. Серверная часть также использует библиотеки для обработки естественного языка и TensorFlow, для применения моделей анализа текста к данным из социальной сети ВКонтакте или введенным пользователем текстам. Серверная часть также поддерживает авторизацию пользователей через ВКонтакте.

Серверная часть состоит из нескольких модулей, которые выполняют различные функции. Например:

- `app.py` - это основной модуль, который запускает сервер и определяет маршруты для обработки запросов клиентов.
- `auth.py` - это модуль, который реализует логику авторизации пользователей через ВКонтакте и получает токен доступа к данным пользователя.
- `vk.py` - это модуль, который реализует логику взаимодействия с API ВКонтакте и получает данные постов и комментариев пользователя по токену и `id`.
- `analyze.py` - это модуль, который реализует логику анализа текста с помощью моделей MBTI и Big Five и вычисляет результаты анализа.

- `templates` - это папка, которая содержит HTML-шаблоны для отображения страниц с формой анализа текста, результатами анализа или ошибками.

- `static` - это папка, которая содержит статические файлы, такие как CSS-стили или изображения.

Клиентская часть — это веб-браузер пользователя, который отправляет HTTP-запросы на серверную часть и получает HTML-страницы с результатами анализа текста. Клиентская часть также позволяет пользователю вводить текст для анализа или авторизоваться через ВКонтакте для получения данных постов и комментариев.

Клиентская часть состоит из нескольких страниц, которые предоставляют различные функции. Например:

- `index.html` - это главная страница приложения, которая содержит форму для ввода текста для анализа или кнопку для авторизации через ВКонтакте.

- `results.html` - это страница с результатами анализа текста. Страница также содержит графики и таблицы для визуализации результатов.

- `error.html` - это страница с сообщением об ошибке в случае невалидного `id` или текста или отсутствия данных.

Внешние сервисы — это API социальной сети ВКонтакте1234, которое предоставляет доступ к данным пользователя и постов по токену и `id`. Внешние сервисы также проверяют права доступа и наличие данных.

Внешние сервисы состоят из нескольких методов API ВКонтакте1234, которые используются для получения данных пользователя и постов. Например:

- `users.get` — это метод API ВКонтакте1234, который возвращает информацию о пользователе по `id`.

- `wall.get` — это метод API ВКонтакте1234, который возвращает список постов на стене пользователя по `id`.

- `wall.getComments` - это метод API ВКонтакте1234, который возвращает список комментариев к посту на стене пользователя по `id`.

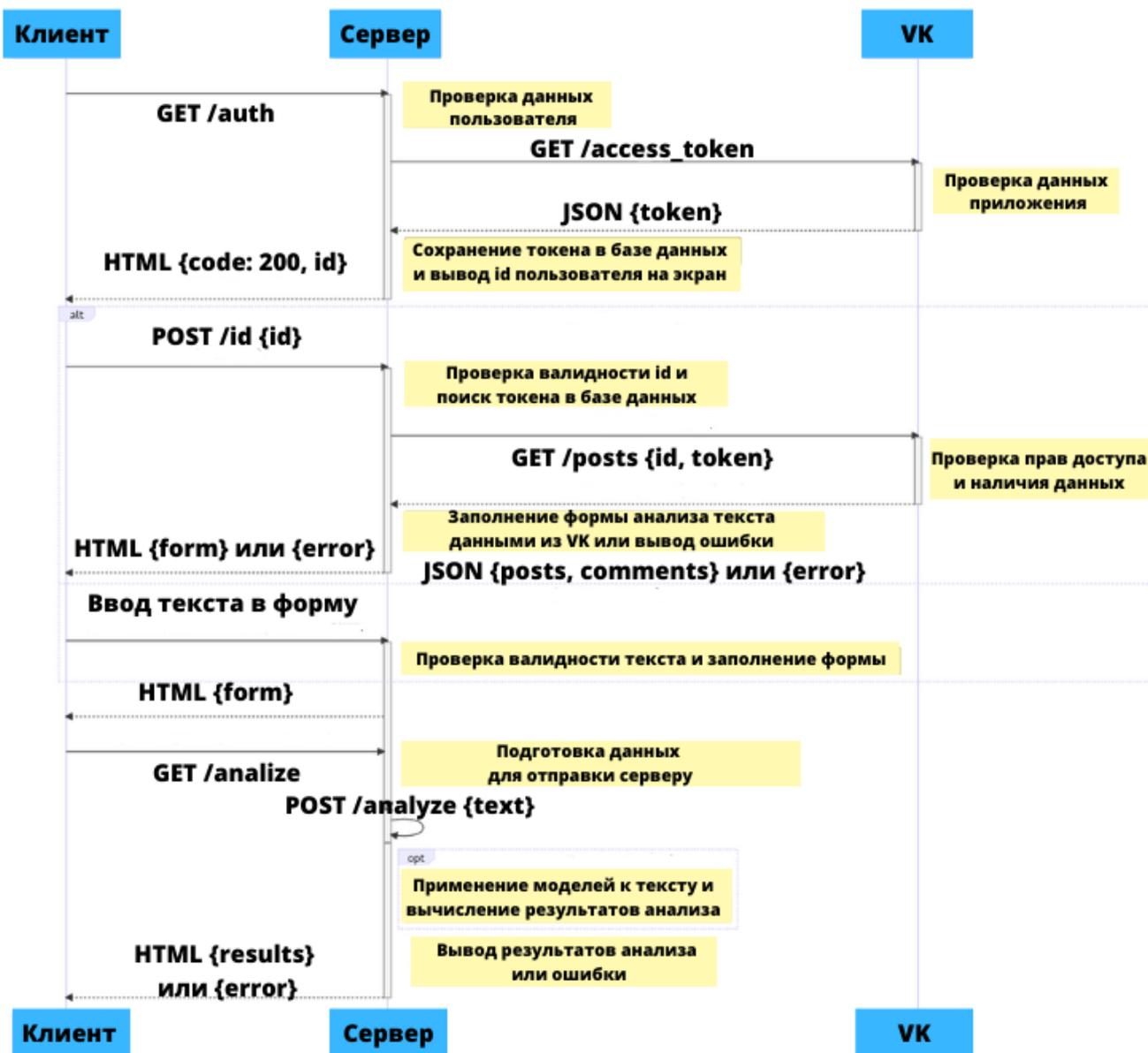


Рисунок 10 - Диаграмма последовательностей

На диаграмме 10 видно, что приложение работает следующим образом:

- Клиент отправляет GET-запрос на сервер для авторизации через ВКонтакте.
- Сервер проверяет данные пользователя и отправляет GET-запрос на ВКонтакте для получения токена доступа к данным пользователя.

- ВКонтакте проверяет данные приложения и возвращает JSON-ответ с токеном или ошибкой.
- Сервер сохраняет токен в базе данных и возвращает HTML-ответ с кодом 200 и id пользователя или ошибкой.
- Клиент может отправить POST-запрос на сервер с id пользователя для получения данных из ВКонтакте или ввести текст для анализа в форму без авторизации.
- Сервер проверяет валидность id или текста и отправляет GET-запрос на ВКонтакте для получения данных постов и комментариев или заполняет форму анализа текста данными из текста без авторизации.
- ВКонтакте проверяет права доступа и наличие данных и возвращает JSON-ответ с данными постов и комментариев или ошибкой.
- Сервер заполняет форму анализа текста данными из ВКонтакте или выводит ошибку.
- Клиент нажимает кнопку “Анализировать текст” и отправляет POST-запрос на сервер с текстом для анализа.
- Сервер подготавливает данные для отправки серверу анализа текста и отправляет POST-запрос на сервер анализа текста с текстом для анализа.
- Сервер анализа текста применяет модели анализа текста к тексту и вычисляет результаты анализа или выводит ошибку.
- Сервер анализа текста возвращает JSON-ответ с результатами анализа или ошибкой.
- Сервер выводит результаты анализа или ошибку на экран в виде HTML-страницы.

1.9.2.2 Итоговые скриншоты приложения

На последующих скриншотах показаны внешний вид и результаты работы приложения.

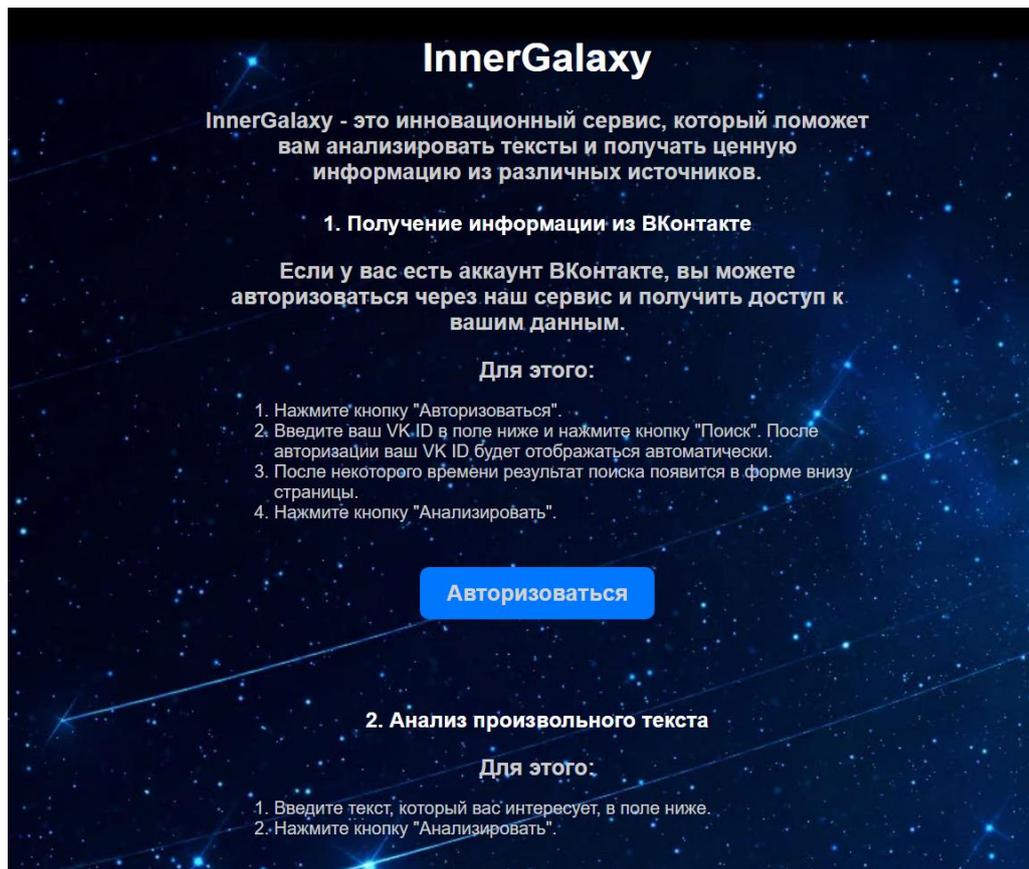


Рисунок 11 – Приветственное окно программы

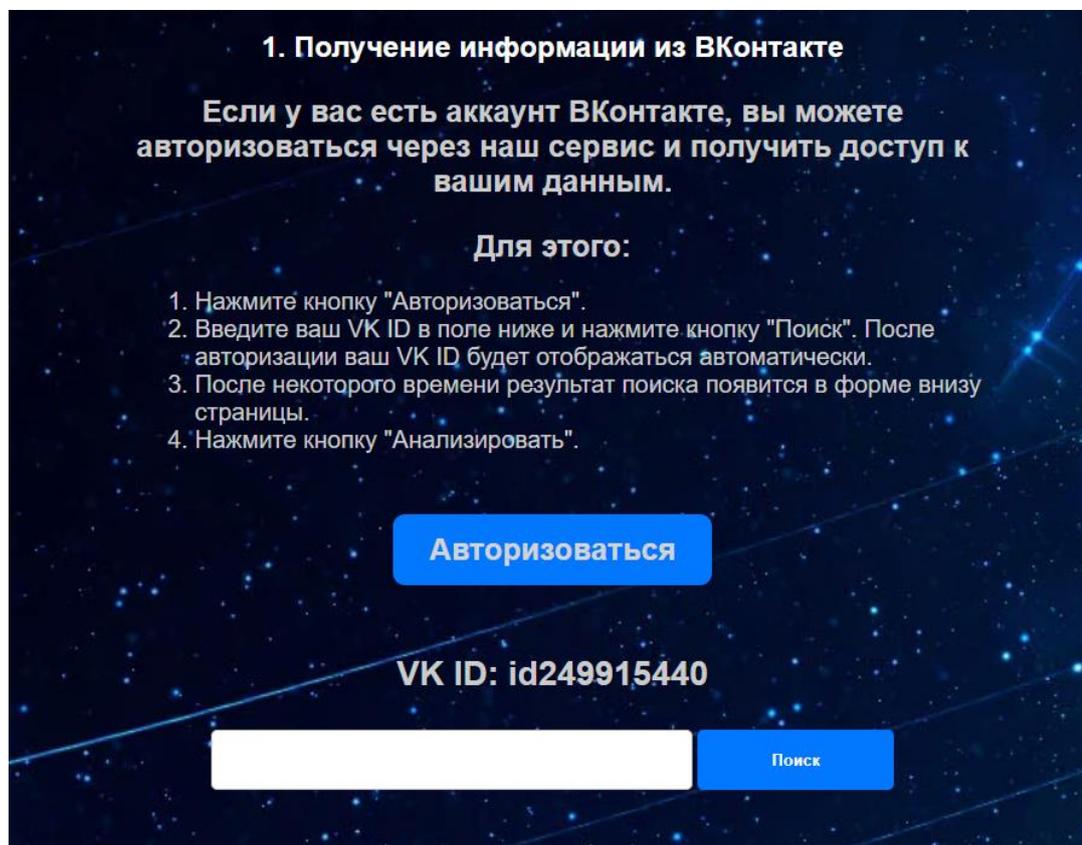


Рисунок 12 – Раздел авторизации ВКонтакте

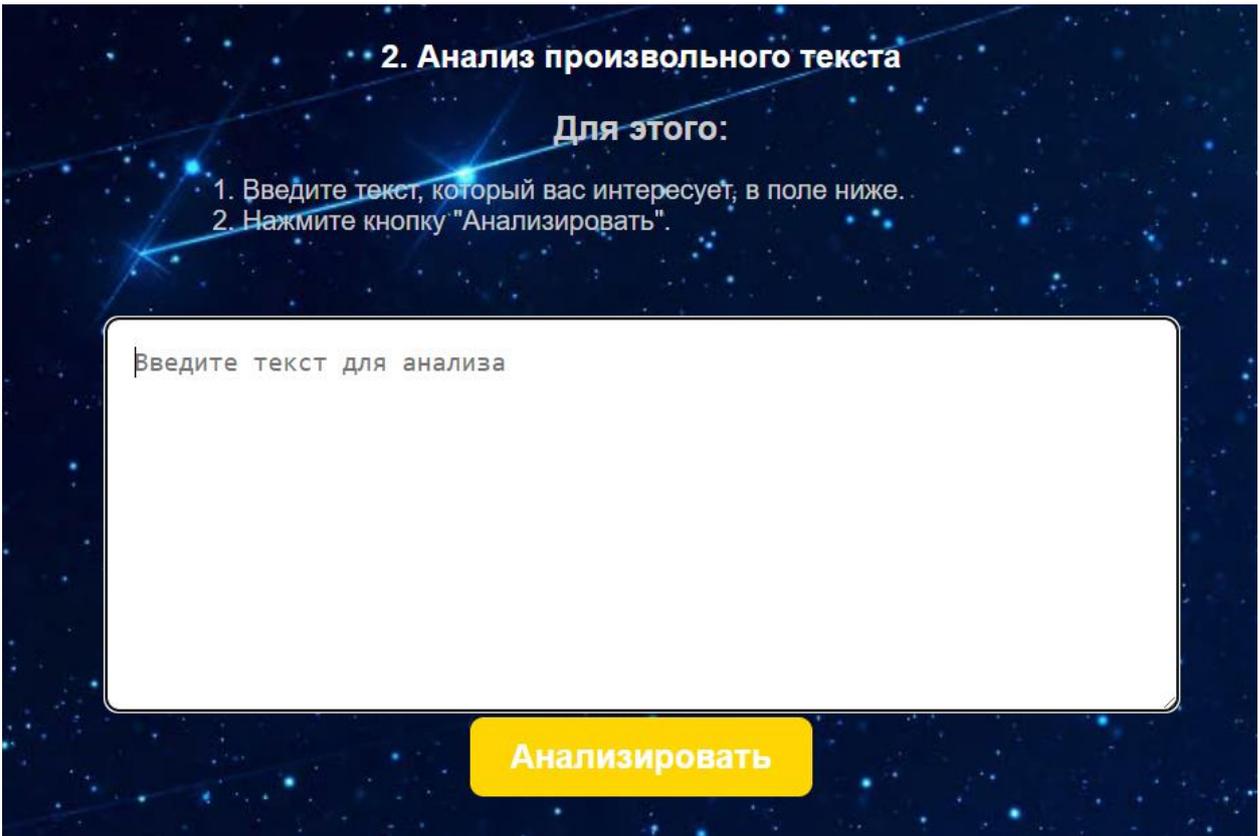


Рисунок 13 – Окно для ввода анализируемого текста

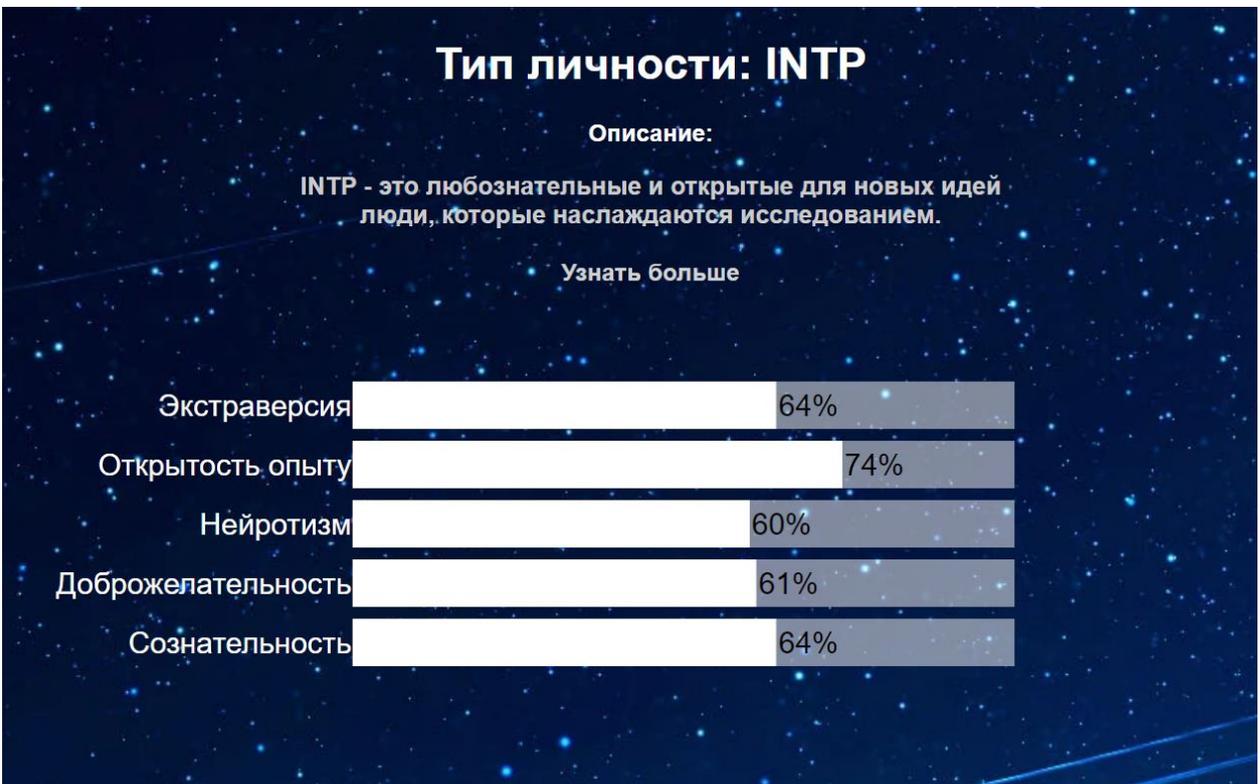


Рисунок 14 – Окно результатов анализа

1.10 Выводы по разделу

В разделе был описан процесс создания веб-приложения для определения типа личности по тексту. Были рассмотрены две типологии личности: Big Five и MBTI, а также разработан ряд моделей на базе DEBERTa для определения различных черт личности с помощью фреймворка TensorFlow. Приложение реализовано с использованием фреймворка Flask и VK API для получения и анализа реальных данных из социальной сети ВКонтакте. Были проведены тесты и оценка качества работы приложения на разных типах текстов. Приложение демонстрирует высокую точность и скорость определения типа личности: средняя точность составляет 65% по шкале MBTI и 75% по шкале Big Five, а среднее время определения типа личности по одному тексту не превышает 4 секунды.

Определение типа личности по тексту актуально для разных сфер жизни и деятельности человека. Тип личности влияет на многие аспекты психики и поведения человека, а также на его взаимодействие с другими людьми. Также определение типа личности по тексту может быть полезно для психологов, коучей, учителей, маркетологов, рекрутеров и бизнесменов, которые могут использовать эту информацию для различных целей, связанных с психологическим консультированием, образованием, маркетингом, рекрутингом и оптимизацией бизнес-процессов.

**ЗАДАНИЕ К РАЗДЕЛУ
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСООБЪЕКТИВНОСТЬ
И РЕСУРСОСБЕРЕЖЕНИЕ»**

Обучающемуся:

Группа	ФИО
8К93	Третьяков Дмитрий Андреевич

Школа	Инженерная школа информационных технологий и робототехники	Отделение школы (НОЦ)	Отделение информационных технологий
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>	Бюджет проекта НИ – 388 767 руб. Затраты на заработную плату – 257 076 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	Премиальный коэффициент 30%; Коэффициент доплат и надбавок 20%; Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Коэффициент отчислений на уплату во внебюджетные фонды 15%, 30%

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	Определение потенциальных потребителей; SWOT-анализ;
2. <i>Планирование и формирование бюджета научных исследований</i>	Формирование плана и графика разработки; Определение структуры и трудоемкости работ; Разработка графика Ганта; Формирование бюджета затрат;
3. <i>Определение ресурсной, финансовой, экономической эффективности</i>	Оценка эффективности исследования

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 по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор ОСГН	Гасанов Магеррам Али оглы	Д.э.н.		

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8К93	Третьяков Дмитрий Андреевич		

2. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Разработка НИ производится группой работников, состоящей из двух человек – руководителя и студента. Данная выпускная квалификационная работа заключается в программной реализации модуля психологической типологизации на основе обработки естественного языка.

Целью раздела «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» является определение перспективности и успешности НИ, оценка его эффективности, уровня возможных рисков, разработка механизма управления и сопровождения конкретных проектных решений на этапе реализации.

Для достижения обозначенной цели необходимо решить следующие задачи:

1. Оценить коммерческий потенциал и перспективность разработки НИ;
2. Осуществить планирование этапов выполнения исследования;
3. Рассчитать бюджет затрат на исследования;
4. Произвести оценку научно-технического уровня исследования и оценку рисков.

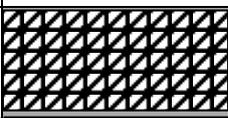
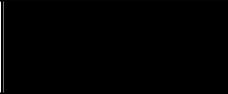
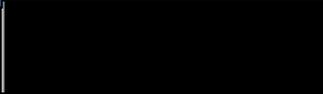
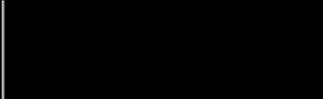
2.1 Потенциальные потребители результатов исследования

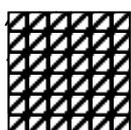
Целевой рынок можно разделить на три сегмента по характеру применения разработки:

- Крупный бизнес, имеющий в своем составе департаменты по подбору персонала, специализированные агентства HR, органы государственной власти, работающие в сфере трудовой занятости населения, а также силовые органы, в компетенции которых входит надзор за медиа пространством и другие потребности в анализе текстовой информации, применяющие ПО на регулярной основе в профессиональной сфере;

- Медицинские организации и практикующие врачи, специализирующиеся на оказании услуг в сфере психологии, могут применять ПО для повышения производительности труда, или в качестве мотивирующих инструментов;
- Частные лица или ИП, применяющие ПО разово или нерегулярно, возможно в качестве развлечения,

Таблица 5 – Карта сегментирования рынка продаж

		Применение		
		Работа с персоналом	Некоммерческое применение	Коммерческое применение
Размер компании	Крупный бизнес			
	Медицина			
	Частные лица			



IBM Watson
Personality Insights



OKOCRМ Зигмунд



Проект ВКР

Согласно карте сегментирования рынка, можно сделать вывод об относительной свободе всех сегментов продаж программных решений по теме данной работы, но основным и приоритетным сегментом будет являться рынок крупного бизнеса. Кроме того, привлекательными являются остальные ранки.

2.2 Анализ конкурентных технических решений

Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения будет основываться на сравнении разрабатываемой системы (1) и двух конкурентных решений производителей программного обеспечения, а именно программного

комплекса «IBM Watson Personality Insights» (2) производство США и отечественной разработки OKOCRM (3). Импортное ПО это инструмент, который извлекает и анализирует спектр личностных качеств, помогая находить полезную информацию о людях и объектах и, в свою очередь, помогает конечным пользователям о высоко персонализированные взаимодействия, является продуктом высочайшего уровня, но в силу текущей геополитической обстановки не доступен для приобретения бизнесом в РФ. OOKOCRM – представляет собой CRM систему для психологов, основной упор в ней сделан на автоматизацию работы с клиентами.

Анализ конкурентных технических решений был проведен с помощью оценочной карты, представленной в таблице 6 (индекс 1 отвечает за текущее решение).

Таблица 6 – оценочная карта сравнения конкурентных технических решений

№	Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
			Бк1	Бк2	Бк3	К1	К2	К3
Технические критерии оценки ресурсоэффективности								
4	Ресурсоэффективность	0,1	5	5	4	0,5	0,5	0,4
5	Надежность	0,1	3	3	5	0,3	0,3	0,5
6	Потребность в ресурсах памяти	0,1	4	4	3	0,4	0,4	0,3
7	Функциональная мощность (предоставляемые возможности)	0,25	1	3	2	0,25	0,75	0,5
8	Простота эксплуатации	0,1	2	2	4	0,2	0,2	0,4
9	Качество пользовательского интерфейса	0,07	1	5	5	0,07	0,35	0,35
Экономические критерии оценки эффективности								
1	Конкурентоспособность продукта	0,05	2	5	4	0,1	0,25	0,2
2	Уровень проникновения на рынок	0,05	1	3	1	0,05	0,15	0,05
3	Цена	0,09	2	1	2	0,18	0,09	0,18
4	Послепродажное обслуживание	0,09	2	4	4	0,18	0,36	0,36

Итого	1	23	45	34	2,23	3,35	3,24
-------	---	----	----	----	------	------	------

Анализ конкурентных решений определяется по формуле 1:

$$K = \sum V_i \times B_i, \quad (1)$$

где K – конкурентоспособность научной разработки или конкурента;

V_i – вес показателя (в долях единицы);

B_i – балл i -го показателя.

В целом разработка уступает индустриальным конкурентам, но разрыв позволяет предположить о возможности продукта конкурировать с ними. Кроме того, по отдельным пунктам: надежность, простота эксплуатации, и цена, разработка не уступает или опережает конкурентов.

2.3 Технология Quad

Технология Quad (Quality ADvistor) представляет собой гибкий инструмент измерения характеристик, описывающий качество новой разработки и ее перспективность на рынке и позволяющие принимать решение целесообразности вложения денежных средств в научно-исследовательский проект.

Анализ имеющихся конкурентных продуктов необходимо проводить с достаточной регулярностью, так как рынок ИТ находится в постоянном и активном движении на сегодняшний день. Данный анализ позволяет производить изменения текущего внедрения, чтобы сделать его наиболее перспективным и конкурентоспособным в дальнейшем. Результат QuaD-анализа представлен в таблице 7.

Таблица 7 – Оценка критериев в соответствии с технологией Quad

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение	Средневзвешенное значение
Ресурсоэффективность	0,1	95	100	0,95	9,5
Надежность	0,1	95	100	0,95	9,5
Потребность в	0,1	85	100	0,85	8,5

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение	Средневзвешенное значение
ресурсах памяти					
Функциональная мощность (предоставляемые возможности)	0,25	70	100	1,75	17,5
Простота эксплуатации	0,1	85	100	0,85	8,5
Качество пользовательского интерфейса	0,07	80	100	0,56	5,6
Конкурентоспособность продукта	0,05	90	100	0,45	4,5
Уровень проникновения на рынок	0,05	70	100	0,35	3,5
Цена	0,09	95	100	0,855	8,55
Послепродажное обслуживание	0,09	80	100	0,72	7,2
Итого:					82,85

Средневзвешенное значение показателя качества и перспективности научной разработки равен 82,85 (попадает в промежуток 80-100), то есть такая разработка считается перспективной.

2.4 SWOT-анализ

SWOT-анализ – метод стратегического планирования, заключающийся в выявлении факторов внутренней и внешней среды организации и разделении их на четыре категории:

- Strengths (сильные стороны);
- Weaknesses (слабые стороны);
- Opportunities (возможности);
- Threats (угрозы).

Сильные (S) и слабые (W) стороны являются факторами внутренней среды объекта анализа, (то есть тем, на что сам объект способен повлиять); возможности (O) и угрозы (T) являются факторами внешней среды (то есть тем, что может повлиять на объект извне и при этом не контролируется

объектом). В рамках данного анализа выявлены сильные и слабые стороны проекта, а также его возможности и угрозы. Все параметры SWOT анализа данного проекта представлены в таблице 8.

Таблица 8 – Сильные и слабые стороны проекта

	Сильные стороны: С1. Низкие издержки. С2. Адаптируемость продуктов, возможность настройки под индивидуальные потребности бизнеса. С3. Относительно низкая цена.	Слабые стороны: Сл1. Средний уровень персонала. Сл2. Отсутствие опыта реализации глобальных проектов. Сл3. Ограниченность материальных ресурсов.
Возможности: В1. Слабая конкуренция на региональных рынках. В2. Льготная система налогообложения		
Угрозы: У1. Возможность появления новых конкурентов. У2. Ужесточение фискальной политики (повышение налоговых ставок, введение новых налогов).		

В рамках второго этапа проведём анализ соответствия сильных и слабых сторон проекта внешним условиям окружающей среды. Анализ представлен ниже в виде интерактивных матриц, сильное соответствие отмечено знаком «+», слабое знаком «-».

Таблица 9 – Интерактивная матрица сильных сторон и возможностей проекта

		Сильные стороны		
		С1	С2	С3
Возможности	В1	+	+	+
	В2	-	-	-

Направления реализации сильных сторон и возможностей: В1С1С3В2С3.

Таблица 10 – Интерактивная матрица слабых сторон и возможностей проекта

Слабые стороны				
Возможности		Сл1	Сл2	Сл3
	B1	+	+	-
	B2	-	0	+

Направления реализации слабых сторон и возможностей: B1Сл1Сл2, B2Сл3.

Таблица 11 – Интерактивная матрица сильных сторон и угроз проекта

Сильные стороны				
Угрозы		С1	С2	С3
	У1	+	+	-
	У2	-	0	+

Направления реализации сильных сторон и угроз: У1С1С2С3, У2У3С3С4, У4С1С2С3.

Таблица 12– Интерактивная матрица слабых сторон и угроз проекта

Слабые стороны				
Угрозы		Сл1	Сл2	Сл3
	У1	-	-	+
	У2	+	-	-

Направления реализации слабых сторон и угроз: У2Сл1, У1Сл3.

В рамках третьего этапа построим итоговую матрицу SWOT анализа, результат представлен в таблице 13.

Таблица 13 – Итоговая матрица SWOT

	Сильные стороны: С1. Низкие издержки. С2. Адаптируемость продуктов, возможность настройки под индивидуальные потребности бизнеса. С3. Относительно низкая цена.	Слабые стороны: Сл1. Средний уровень персонала. Сл2. Отсутствие опыта реализации глобальных проектов. Сл3. Ограниченность материальных ресурсов.
Возможности: В1. Слабая конкуренция на региональных рынках. В2. Льготная система налогообложения	Внедрение ПО в РФ и за рубежом, придерживаться оптимальной ценовой политики	Привлечение специалистов на постоянной основе и разовые работы, обучение персонала, совместные работы с другими организациями, заимствование опыта.
Угрозы: У1. Возможность появления новых конкурентов. У2. Ужесточение фискальной политики (повышение налоговых ставок,	Наращивание функционала и качества ПО	Привлечение финансирования и использования гос. программ по импортозамещению в сфере ПО

введение налогов).	НОВЫХ		
-----------------------	-------	--	--

SWOT-анализ показал, что риски, связанные возможно устранить расширением функционала ПО, расширением рынков сбыта, работой с персоналом и использованием программ господдержки. В целом, несмотря на угрозы и слабые стороны проекта, разработка обладает является перспективной.

2.5 Планирование научно-исследовательских работ

2.5.1 Структура работ в рамках научного исследования

Важным этапом проведения научно-исследовательских работ является необходимость планирования работ, которое включает в себя определение полного перечня работ, а также их распределение между всеми исполнителями проекта. Исполнителями проекта являются студент и научный руководитель. Научный руководитель определяет цели и задачи для студента, направляет и контролирует его работу, оценивает результаты проделанной работы и дает рекомендации студенту. Студент полностью отвечает за выполняемую работу. В таблице 14 представлен перечень работ, а также распределение исполнителей по ним.

Таблица 14 - Перечень работ и распределение исполнителей

№	Название	Исполнитель
1	Получение индивидуального задания	Научный руководитель, Студент

№	Название	Исполнитель
2	Планирование работ	Научный руководитель, Студент
3	Изучение литературных источников	Студент
4	Изучение технологий	Студент
5	Разработка алгоритмов моделирования	Студент
6	Проектирование программы	Студент
7	Проектирование интерфейса программы	Студент
8	Кодирование программы	Студент
9	Проектирование входных данных, загрузка данных	Студент
10	Первичная подготовка данных	Студент
11	Проектирование модуля доступа к данным	Студент
12	Проектирование интерфейса модуля доступа к данным	Студент
13	Кодирование модуля доступа к данным	Студент
14	Объединение модулей	Студент
15	Тестирование	Студент
16	Согласование выполненной работы с научным руководителем	Научный руководитель, Студент
17	Документирование разработки	Студент
18	Выполнение разделов социальная ответственность и финансовый менеджмент	Студент
19	Оформление работы	Студент

2.5.2 Определение трудоемкости выполнения работ

Как правило, трудовые затраты образуют основную часть стоимости исследования, поэтому важным этапом планирования научно-исследовательской деятельности является определение трудоемкости работ.

Определение трудоёмкости выполнения работ осуществляется на основе экспертной оценки ожидаемой трудоёмкости выполнения каждой работы путём расчёта длительности работ в рабочих и календарных днях каждого этапа работ. Трудоемкость оценивается по следующей формуле 2:

$$t_{ож\ i} = \frac{3t_{mini} + 2t_{maxi}}{5}, \quad (2)$$

где $t_{ож\ i}$ – это ожидаемая трудоемкость i -ой работы (чел.-дни),

t_{mini} – это минимально возможная трудоемкость выполнения заданной i -ой работы (оптимистическая оценка) (чел.-дни),

t_{maxi} – это максимально возможная трудоемкость выполнения заданной i -ой работы (пессимистическая оценка) (чел.-дни).

После оценки ожидаемой трудоемкости работ, производится определение продолжительности каждой работы в рабочих днях по формуле 3:

$$Tpi = \frac{t_{ож\ i}}{ч\ i}, \quad (3)$$

где Tpi – это продолжительность одной работы (раб. дни),

$t_{ож\ i}$ – это ожидаемая трудоемкость выполнения одной работы (чел.-дни),

$ч\ i$ – это численность исполнителей, выполняющих одновременно одну и ту же работу на этом этапе (чел.).

Для того чтобы в дальнейшем построить график работ с помощью диаграммы Ганта, необходимо также произвести перевод длительности работ из рабочих дней в календарные по формуле 4:

$$Tki = Tpi \times k_{кал}, \quad (4)$$

где Tki – это продолжительность выполнения i -й работы в календарных

днях;

T_{pi} – это продолжительность выполнения i -й работы в рабочих днях;

$k_{\text{кал}}$ – это коэффициент календарности.

Коэффициент календарности $k_{\text{кал}}$ рассчитывается по формуле 5:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}, \quad (5)$$

где $k_{\text{кал}}$ – коэффициент календарности;

$T_{\text{кал}}$ – количество календарных дней в году;

$T_{\text{кал}}$ – количество выходных дней в году;

$T_{\text{пр}}$ – количество праздничных дней в году.

С учётом того, что календарных дней в 2023 году 365, а сумма выходных и праздничных дней составляет 118 дней, коэффициент календарности равен $k_{\text{кал}} = 1,48$.

2.5.3 Разработка графика проведения научного исследования

В таблице 15 приведены временные показатели научного исследования.

Таблица 15 - Временные показатели проведения научного исследования

№	Название	Исполнитель	Трудоёмкость работ, чел-дни		Длительность работ, дни		
			tmin	tmax	тож	Тр	Тк
1	Получение индивидуального задания	Научный руководитель, Студент	2	5	3,2	1,6	2,4
2	Планирование работ	Научный руководитель, Студент	2	5	3,2	1,6	2,4

№	Название	Исполнитель	Трудоемкость работ, чел-дни		Длительность работ, дни		
			tmin	tmax	тож	Тр	Тк
3	Изучение литературных источников	Студент	6	9	7,2	7,2	10,7
4	Изучение технологий	Студент	6	10	7,6	7,6	11,2
5	Разработка алгоритмов моделирования	Студент	5	8	6,2	6,2	9,2
6	Проектирование программы	Студент	5	10	7,0	7,0	10,4
7	Проектирование интерфейса программы	Студент	3	7	4,6	4,6	6,8
8	Кодирование программы	Студент	8	12	9,6	9,6	14,2
9	Проектирование входных данных, загрузка данных	Студент	3	5	3,8	3,8	5,6
10	Первичная подготовка данных	Студент	4	6	4,8	4,8	7,1
11	Проектирование модуля доступа к данным	Студент	3	7	4,6	4,6	6,8
12	Проектирование интерфейса модуля доступа к данным	Студент	3	7	4,6	4,6	6,8
13	Кодирование модуля доступа к данным	Студент	3	5	3,8	3,8	5,6
14	Объединение модулей	Студент	3	5	3,8	3,8	5,6
15	Тестирование	Студент	2	4	2,8	2,8	4,1
16	Согласование выполненной работы с научным руководителем	Научный руководитель, Студент	2	4	2,8	1,4	2,1
17	Документирование разработки	Студент	2	5	3,2	3,2	4,7

№	Название	Исполнитель	Трудоемкость работ, чел-дни		Длительность работ, дни		
			tmin	tmax	тож	Тр	Тк
18	Выполнение разделов социальная ответственность и финансовый менеджмент	Студент	5	10	7,0	7,0	10,4
19	Оформление работы	Студент	5	10	7,0	7,0	10,4
	Итого	Научный руководитель	6	14	9,2	9,2	13,6
		Студент	72	134	96,8	96,8	143,3

На основе данных таблицы 15 был построен календарный план-график, представленный на рисунке 15 .

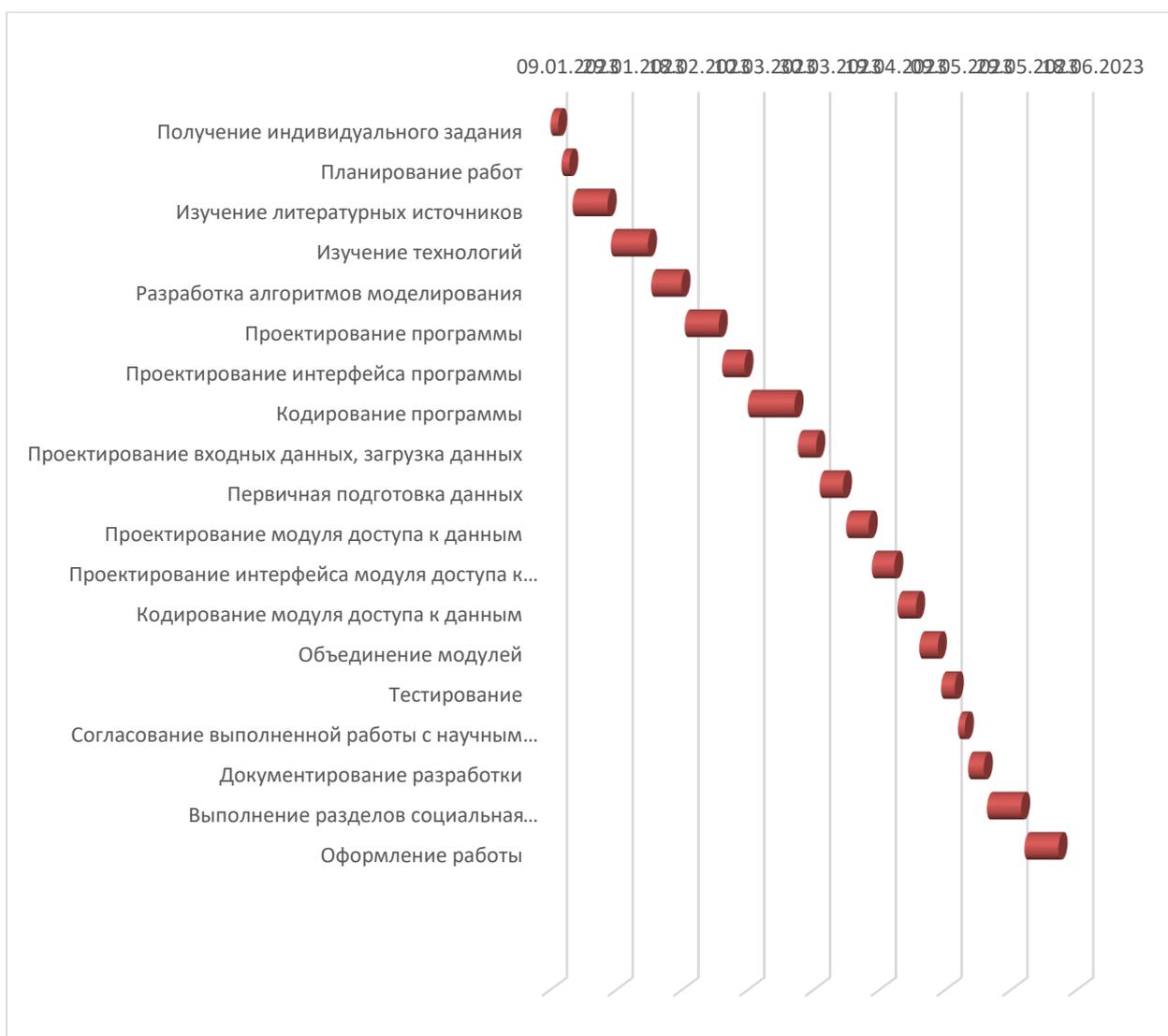


Рисунок 15 – Календарный график проведения работ

2.6 Бюджет научно-исследовательских работ

В состав бюджета входит стоимость всех расходов, необходимых для выполнения работ по проекту. При формировании бюджета используется группировка затрат по следующим статьям:

- расчет материальных затрат НТИ;
- затраты на специальное оборудование;
- основная заработная плата исполнителей;
- дополнительная заработная плата исполнителей;
- отчисления во внебюджетные фонды;

- накладные расходы.

2.6.1 Расчет материальных затрат НИР

Материальные затраты на создание рабочих мест научного руководителя и студента, включающих в себя комплект офисной мебели, в текущем проекте отсутствуют, так как они уже существуют и нет необходимости в их дополнении или модернизации под текущий проект. Амортизационные отчисления также отсутствуют по причине истекших (более шести лет) сроков уплаты отчислений. Статья аренды помещений не включена, так как разработка проекта осуществляется в дистанционном режиме. Затраты на канцелярские принадлежности будем учитывать, как накладные расходы.

Рассчитаем других статей материальных затрат НИР. Стоимость одного дня затрат электроэнергии рассчитывается по формуле 6.

$$\text{Электр} = \text{Эл мощн} \times \text{Эл тариф} \times \text{Эл р.день} \quad (6)$$

где Эл мощн – потребляемая электрическая мощность, КВтч

Эл тариф – Тариф для населения, руб/КВтч

Эл р.день – Количество часов в рабочем дне, час

Таким образом расчет затрат за рабочий день составит

$$\text{Электр} = 0,300 \times 4,39 \times 8 = 10,54 \text{ рублей}$$

Стоимость затрат на услуги связи принимаем равный тарифному плану «100 навсегда» провайдера «Зеленая точка» 380 руб/месяц.

Материальные затраты НИР представлены в таблице 16.

Таблица 16 - Материальные затраты

Наименование статьи расходов	Единица измерения	Количество	Цена за единицу, руб	Стоимость, руб
Компенсация затрат на электроэнергию	день	107	10,54	1127,78

Наименование статьи расходов	Единица измерения	Количество	Цена за единицу, руб	Стоимость, руб
Компенсация затрат на услуги связи	месяц	4	380	1520
Итого				2647,78

Таким образом, сумма материальных затрат НИР составляет 2647,78 рублей.

2.6.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей

Данная статья затрат включает в себя затраты на приобретение специального оборудования. Также в эту статью включаются затраты по доставке и монтажу оборудования, равные 15% от его стоимости.

В ходе работы над проектом использовалось оборудование, имеющееся у исполнителей, соответственно необходим расчет его амортизации.

При создании информационной системы был использован один ноутбук и один персональный компьютер общей стоимостью 94000 рублей. Стоимость программного обеспечения ОС Microsoft Windows 10, Microsoft Office 2016, приобреталась совместно с ноутбуком и персональным компьютером составляет 6000 рублей. Остальное программное обеспечение (Microsoft Visual Studio Code, Nvidia Deep Learning Software, Python & etc) в проекте используется по GPU лицензиям, затраты отсутствуют.

Расчет амортизации: первоначальная стоимость 100000 рублей; срок полезного использования для машин офисных код 330.26.20.11 составляет 36 месяцев. Планируемое время использования для написания ВКР - 6 месяцев. Норма амортизация основных средств линейным способом рассчитывается по формуле 7:

$$A_n = \frac{1}{n} \times 100\%, \quad (7)$$

n – установленный срок в месяцах;

A_n – норма амортизации.

Тогда расчет амортизации ПК:

Норма амортизации:

$$A_n = \frac{1}{n} \times 100\% = \frac{1}{36} \times 100\% = 2,78\%$$

Ежемесячные амортизационные отчисления:

$$A_m = 100000 \times 0,0278 = 2780 \text{ рублей}$$

Итоговая сумма амортизации основных средств:

$$A = 2780 \times 8 = 22240 \text{ рублей}$$

Таким образом, сумма затрат на специальное оборудование составляет 22240 рублей, в виде амортизационных отчислений.

2.6.3 Основная заработная плата исполнителя темы

Данная статья затрат включает основную заработную плату, премии и доплаты всех исполнителей проекта. В качестве исполнителей проекта выступают студент и научный руководитель. Заработная плата рассчитывается по формуле 8:

$$Z_{зп} = Z_{осн} + Z_{доп}, \quad (8)$$

где $Z_{зп}$ – заработная плата исполнителя;

$Z_{осн}$ – основная заработная плата исполнителя;

$Z_{доп}$ – дополнительная заработная плата исполнителя (12%-15% от размера основной заработной платы).

Основную заработную плату можно получить по формуле 9:

$$Z_{осн} = Z_{дн} \times T_p \times (1 + (K_{пр} + K_{ДФ})) \times K_r \quad (9)$$

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

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

$K_{ДФ}$ – коэффициент доплат и надбавок (0,2-0,5);

K_r – районный коэффициент (для Томска 1,3);

Тр – продолжительность работ, выполняемых работником, раб. дни

где Здн – среднедневная заработная плата исполнителя;

Тр – продолжительность работ, выполняемых исполнителем.

Среднедневную заработную плату можно получить по формуле 10:

$$\text{Здн} = \frac{\text{Зм} \times \text{М}}{\text{Фд}}, \quad (10)$$

где Зм – месячный должностной оклад исполнителя, рубли;

М – количество месяцев работы равно:

При отпуске в 24 рабочих дня М = 11,2 месяца, 5-дневная неделя;

При отпуске в 48 рабочих дней М = 10,4 месяца, 6-дневная неделя;

Фд – действительный годовой фонд рабочего времени персонала по разработке.

Должностные оклады исполнителей проекта представлены в таблице 17.

Таблица 17 - Месячные должностные оклады исполнителей

Исполнитель	Районный коэффициент (для Томска)	Размер месячного оклада без учета коэффициента, рубли
Научный руководитель (должность-доцент, степень – кандидат технических наук)	1,3	33664
Студент	1,3	25000

Баланс рабочего времени для 6-дневной рабочей недели представлен в таблице 18.

Таблица 18 - Баланс рабочего времени (для 6-дневной недели)

Показатели рабочего времени	Дни
Календарные дни	365
Нерабочие дни (праздники/выходные)	118
Действительный годовой фонд рабочего времени	247

На основе формулы 8 и таблиц 16-17 была рассчитана среднедневная заработная плата:

$$\text{Здн(студент)} = \frac{25000 \times 10,4}{244} = 1065,57 \text{ руб.}$$

$$\text{Здн(науч. рук.)} = \frac{33664 \times 10,4}{243} = 1434,86 \text{ руб.}$$

Расчет затрат на основную заработную плату приведен в таблице 19.

Таблица 19 - Затраты на основную заработную плату

Исполнители	Здн, руб.	Кпр	Кд	Кр	Тр	Зосн,руб.
Научный руководитель	1 434,86 Р	0,3	0,2	1,3	10	27 979,75 Р
Студент	1 065,57 Р	0,3	0,2	1,3	97	201 553,28 Р
Итого:						229 533,03 Р

Итоговая сумма затрат на основную заработную плату составила 229 533,03 руб.

2.6.4 Расчет дополнительной заработной платы исполнителей темы

Данная статья расходов учитывает величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда и выплат, связанных с обеспечением гарантий и компенсаций. Расчёт дополнительной заработной платы осуществляется по формуле 11:

$$\text{Здоп} = k_{\text{доп}} \times \text{Зосн}, \quad (11)$$

где Здоп – дополнительная заработная плата, рубли;

кдоп – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,2);

Зосн – основная заработная плата, рубли.

Расчет затрат на дополнительную заработную плату приведен в таблице 20.

Таблица 20 - Затраты на дополнительную заработную плату

Исполнители	Зосн, руб.	Кдоп	Зосн,руб.
Научный руководитель	27 979,75 Р	0,12	3 357,57 Р
Студент	201 553,28 Р	0,12	24 186,39 Р
Итого:			27 543,96 Р

Итоговая сумма затрат на дополнительную заработную плату составила 27 543,96 руб.

2.6.5 Отчисления во внебюджетные фонды (страховые отчисления)

Федеральным законом от 14.07.22 г. № 239-ФЗ введен единый тариф страховых взносов с 2023 года в размере 30%.

К отчислениям во внебюджетные фонды относятся отчисления:

- отчисления органам государственного социального страхования (ФСС);
- отчисления в пенсионный фонд (ПФ);
- отчисления медицинского страхования (ФФОМС).

Сумма отчислений во внебюджетные фонды рассчитывается на основе затрат на оплату труда исполнителей и может быть вычислена по формуле 12.

$$Звнеб = kвнеб \times (Зосн + Здоп), \quad (12)$$

где $kвнеб$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и др.);

Однако, в 2023 году на учреждения высшего образования распространяется пониженная ставка тарифа страховых взносов, а именно, для плательщиков перечисленных в пункте 1 статьи 427 Налогового кодекса РФ (для хозяйственных обществ и хозяйственных партнерств, деятельность которых заключается в практическом применении (внедрении) результатов интеллектуальной деятельности (программ для электронных вычислительных машин, баз данных, изобретений, полезных моделей, промышленных образцов, селекционных достижений, топологий интегральных микросхем, секретов производства (ноу-хау), исключительные права на которые принадлежат учредителям, участникам (в том числе совместно с другими лицами) таких хозяйственных обществ, участникам таких хозяйственных партнерств - бюджетным научным учреждениям и автономным научным учреждениям либо образовательным организациям высшего образования, являющимся бюджетными учреждениями, автономными учреждениями) применяется пониженный тариф страховых взносов в размере 15% в

отношении части выплат в пользу физического лица, определяемого по итогам каждого календарного месяца как превышение над величиной минимального размера оплаты труда, установленного федеральным законом на начало расчетного периода (п.2.4 ст. 427 НК РФ). с 01 января 2023 тариф с выплат выше минимального размера оплаты труда (далее - МРОТ) составляет - 15% (размер МРОТ установленный Федеральным законом от 19.12.2022 № 522-ФЗ на 2023 год составляет - 16242 рублей).

Таким образом, отчисления страховых взносов будем проводить по формуле (13) для каждого сотрудника

$$\text{Здн} = ((\text{Зосн} + \text{Здоп}) - (\text{МРОТ} * \text{Мес})) * 0,15 + (\text{МРОТ} * 0,3 * \text{Мес}), \quad (13)$$

Расчет затрат на отчисления во внебюджетные фонды приведен в таблице 21.

Таблица 21 - Отчисления во внебюджетные фонды

Исполнители	Зосн, руб.	Здоп,руб.	Отработанное время, месяц	МРОТ	Звнеб,руб.
Научный руководитель	27 979,75 Р	3 357,57 Р	1	16242	7 136,90 Р
Студент	201 553,28 Р	24 186,39 Р	5	16242	46 042,45 Р
					53 179,35 Р

Итоговая сумма отчислений во внебюджетные фонды составила 53 179,35 руб. Справочно, экономия относительно полного тарифа страховых взносов составила 23 943,75 руб.

2.6.6 Накладные расходы

Накладные расходы – расходы на организацию, управление и обслуживание процесса производства товара, оказания услуг; носят комплексный характер. Накладные расходы вычисляются по формуле 14:

$$\text{Знкл} = \text{сумма статей}(1 - 5) \times \text{кнр}, \quad (14)$$

где кнр – коэффициент накладных расходов (16% от суммы затрат, подсчитанных выше).

Суммы затрат возьмем из предыдущих разделов, суммируем,

полученный подитог умножаем на коэффициент 0.16, результаты расчета приведены в таблице 22.

Таблица 22 - Расчет накладных расходов

Статьи затрат	Сумма, руб.
Материальные затраты НТИ	2 647,78 Р
Затраты на специальное оборудование	22 240,00 Р
Затраты на основную заработную плату	229 533,03 Р
Затраты на дополнительную заработную плату исполнителям проекта	27 543,96 Р
Затраты на отчисления во внебюджетные фонды	53 179,35 Р
Под итог	335 144,12 Р
Коэффициент накладных расходов	0,16
Накладные расходы	53 623,06 Р

Итоговая сумма накладных расходов составила 53 199,41 руб.

2.6.7 Формирование бюджета затрат научно-исследовательского проекта

После того, как была подсчитана каждая из статей расходов, можно приступить к формированию общего бюджета затрат проекта. Итоговый бюджет затрат представлен в таблице 23.

Таблица 23 - Расчет бюджета затрат проекта

Наименование статьи	Сумма, руб.	Удельный вес, %
Материальные затраты	2 647,78 Р	0,68
Затраты на специальное оборудование	22 240,00 Р	5,72
Затраты на основную заработную плату	229 533,03 Р	59,04

Наименование статьи	Сумма, руб.	Удельный вес, %
Затраты на дополнительную заработную плату	27 543,96 Р	7,08
Страховые взносы	53 179,35 Р	13,68
Накладные расходы	53 623,06 Р	13,79
Общий бюджет	388 767,18 Р	100,00

Таким образом, общий бюджет НИ составляет 388 767,18 руб.

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

Интегральный показатель финансовой эффективности получают в ходе оценки бюджета затрат нескольких вариантов выполнения исследования по формуле 15:

$$I_{\text{фип рипс } i} = \Phi_{\text{ри}} \Phi_{\text{тах}}, \quad (15)$$

где $I_{\text{фип рипс } i}$ – интегральный финансовый показатель разработки;

$\Phi_{\text{ри}}$ – стоимость i -го варианта исполнения;

$\Phi_{\text{тах}}$ – максимальная стоимость исполнения научно - исследовательского проекта.

Интегральный показатель ресурсоэффективности может быть вычислен по следующей формуле 16:

$$I_{\text{ри}} = \sum a_i * b_i, \quad (16)$$

где $I_{\text{ри}}$ – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i – оценка i -го варианта исполнения разработки, выраженная в баллах, устанавливается экспертным путем по выбранной шкале оценивания; n – число параметров сравнения

Определение интегральных показателей эффективности проведем при двух варианта исполнения системы:

1. Научное исследование реализуется в среде разработки Visual Studio Code.

2. Научное исследование реализуется в среде разработки PyCharm.

Существенным различием (и выгодным отличием первого варианта) является оптимизированный пользовательский интерфейс современного редактора с богатой поддержкой кода и навигацией, а также встроенную функцию отладки - без необходимости использования полной IDE, в то время как второй вариант включая интеграцию управления версиями, развертывание, интеграцию с системами отслеживания ошибок и удаленную отладку, но имеет платную лицензию стоимость порядка 20000 рублей.

Расчеты интегрального финансового показателя и интегральных показателей ресурсоэффективности обоих вариантов исполнения приведен в таблице 24 и 25.

Таблица 24 – Расчёт интегрального финансового показателя

Вариант	Φp_i , руб	Φmax , руб	$I_{фип}$ рисп i
Visual Studio Code	388 767,18 Р	408 767,18 Р	0,95
PyCharm	408 767,18 Р		1,00

Таблица 25 - Сравнительная оценка характеристик вариантов исполнения проекта

	Весовой коэффициент параметра	Visual Studio Code	PyCharm
Гибкость настройки	0,10	5,00	4,00
Надежность алгоритмов	0,15	5,00	5,00
Быстродействие	0,25	4,00	3,50
Удобство	0,20	3,00	4,00
Функциональность	0,15	4,00	5,00
Интерфейс	0,15	3,00	2,00
Итого	1,00	3,90	3,88

Интегральный показатель эффективности вариантов исполнения разработки ($I_{испi}$) определяется на основании интегрального показателя

ресурсоэффективности и интегрального финансового показателя по формуле (17):

$$I_{\text{сп}i} = \frac{I_{\text{р}i}}{nI_{\text{фип р}i}}, \quad (17)$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных.

Расчет приведен в таблице 26.

Таблица 26 – Сравнительная эффективность разработки

	Интегральный показатель финансовой эффективности	Интегральный показатель ресурсоэффективности	Интегральный показатель эффективности
Visual Studio Code	0,95	3,90	4,10
PyCharm	1,00	3,88	3,88

Сравнив значения интегральных показателей эффективности, можно сделать вывод, что реализация программного обеспечения в первом исполнении является более эффективным вариантом решения задачи, поставленной в данной работе с позиции финансовой и ресурсной эффективности.

2.8 Выводы по разделу

В рамках данного раздела был проведен комплексный SWOT-анализ проекта, который позволил выявить его сильные и слабые стороны, а также определить соответствие его характеристик внешним факторам. Более того, были выявлены возможные пути дальнейшего развития и совершенствования проекта для повышения его конкурентоспособности. Также была произведена оценка качества и перспективности данного проекта.

Ещё одним ключевым моментом, рассмотренным в данном разделе, является планирование работ, выполняемых в рамках проекта. Составленный

список необходимых работ с распределением исполнителей, а также вычисленная трудоемкость, позволили построить наглядный план-график работ в виде диаграммы Ганта. Это позволило рационально распорядиться временными ресурсами, отведенными на выполнение проекта.

Кроме того, был определен бюджет проекта. Определены затраты на заработные платы исполнителям с учетом страховых отчислений, рассчитаны накладные расходы, а также амортизация оборудования, задействованного в процессе выполнения проекта.

Сравнение интегральных показателей эффективности вариантов исполнения показывает, что наиболее выгодным с позиции финансовой и ресурсной эффективности является вариант исполнения, который и был реализован.

ЗАДАНИЕ К РАЗДЕЛУ «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Обучающемуся:

Группа		ФИО	
8К93		Третьяков Дмитрий Андреевич	
Школа	Инженерная школа информационных технологий и робототехники	Отделение (НОЦ)	Отделение информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	09.03.04 Программная инженерия

Тема ВКР:

<i>Разработка программного модуля психологической типологизации на основе обработки естественного языка</i>	
Исходные данные к разделу «Социальная ответственность»:	
<p>Введение</p> <ul style="list-style-type: none"> – Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика) и области его применения. – Описание рабочей зоны (рабочего места) при разработке проектного решения 	<p>Объект исследования: программный модуль психологической типологизации на основе обработки естественного языка;</p> <p>Область применения: информационные технологии;</p> <p>Рабочая зона: офис;</p> <p>Размеры помещения: 4*6 м;</p> <p>Количество и наименование оборудования рабочей зоны: персональный компьютер - 2 шт.;</p> <p>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне: написание программного кода.</p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<p>1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения:</p> <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<p>ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.</p> <p>ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация»</p> <p>ГОСТ Р 50923-96 – не указан в задании «Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения»</p> <p>ГОСТ 21889-76 Система «человек-машина». Кресло человека-оператора. Общие эргономические требования.</p> <p>СП 2.2.3670-20 «Санитарно-эпидемиологические требования к условиям труда»</p> <p>Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ</p> <p>Федерального закона от 18.06.2017 N 125-ФЗ</p>
<p>2. Производственная безопасность при разработке проектного решения:</p> <ul style="list-style-type: none"> – Анализ выявленных вредных и опасных производственных факторов 	<p>Вредные факторы:</p> <p>статические физические перегрузки, связанные с рабочей позой;</p> <p>монотонный режим работы;</p> <p>отсутствие или недостаток освещённости рабочей зоны;</p> <p>отклонение показателей микроклимата помещения;</p> <p>Опасные факторы:</p> <p>опасность поражения электрическим током;</p> <p>Требуемые средства коллективной и индивидуальной</p>

	защиты от выявленных факторов: устройства для вентиляции и очистки воздуха, источники света, изоляция электрических проводов, устройства защитного заземления.
3. Экологическая безопасность при разработке проектного решения	Воздействие на селитебную зону: утилизация компьютеров, их составляющих, компьютерных аксессуаров и других аппаратных средств; Воздействие на литосферу: образование отходов при поломке оргтехники и утилизации ее составных частей; Воздействие на гидросферу: загрязнение при производстве/утилизации компьютерной техники; Воздействие на атмосферу: выбросы углекислого газа при производстве компьютерной техники.
4. Безопасность в чрезвычайных ситуациях при разработке проектного решения	Возможные ЧС: ураган, землетрясение, наводнение, эпидемия, пожар; Наиболее типичная ЧС: пожар.
Дата выдачи задания для раздела по линейному графику	

Задание выдал консультант по разделу «Социальная ответственность»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Мезенцева Ирина Леонидовна			

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8К93	Третьяков Дмитрий Андреевич		

3. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

3.1 Введение

Данная информационная система предназначена для автоматизированного анализа психологических черт личности по методике MBTI и Big Five на основе анализа текстовых данных.

Областью применения являются любые организации, имеющие подразделения по рекрутингу персонала, маркетинговые подразделения фирм торговли и так далее. Разрабатываемое программное обеспечение позволит провести анализ целевой аудитории и отдельных персон, выдать первичную оценку типа личности, после чего уже можно будет принимать управленческие и операционные решения.

Рабочим местом, как правило, является офисное помещение размером 20-25 м² с одним или несколькими компьютерами, соединенными между собой в локальную вычислительную сеть и имеющими доступ в сеть Интернет. В рабочий процесс входит разработка основного модуля, создание анализируемого массива данных, создание веб-приложения, программная реализация, тестирование комплексной работы архитектуры и его доработки. Сам трудовой процесс характеризуется умственным напряжением, зрительным утомлением, повышенным уровнем концентрации, долгим положением сидя за компьютерным столом, что может привести к нервному напряжению, стрессу, болям в спине и общему дискомфорту.

3.2 Правовые и организационные вопросы обеспечения безопасности

3.2.1 Правовые нормы трудового законодательства РФ

Трудовые отношения между работодателем и работником регулируются посредством «Трудового кодекса Российской Федерации» от 30.12.2001 N 197-ФЗ (ред. от 19.12.2022) [22]. В нём регламентируются права и обязанности работодателя и работника, вопросы организации труда, режим труда и отдыха, оплата и нормирование труда, компенсации работникам, защита персональных данных работника, урегулирование трудовых споров.

Наиболее важные положения относительно трудового распорядка работников:

- Нормальная продолжительность рабочего времени не должна превышать 40 часов в неделю.
- В течение рабочего дня (смены) работнику должен быть предоставлен перерыв для отдыха и питания продолжительностью не более двух часов и не менее 30 минут, который в рабочее время не включается.
- Всем работникам предоставляются выходные дни (еженедельный непрерывный отдых)

Правилами внутреннего трудового распорядка или трудовым договором может быть предусмотрено, что указанный перерыв может не предоставляться работнику, если установленная для него продолжительность ежедневной работы (смены) не превышает четырех часов (в ред. Федерального закона от 18.06.2017 N 125-ФЗ).

3.2.2 Эргономические требования к правильному расположению и компоновке рабочей зоны

Основными элементами рабочего места программиста являются: рабочий стол, рабочий стул (кресло), дисплей, клавиатура, мышь; вспомогательными – пюпитр, подставка для ног.

Согласно ГОСТ Р 50923-96 «Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы

измерения» [23], рабочее место с дисплеем должно обеспечивать оператору возможность удобного выполнения работ в положении сидя и не создавать перегрузки костно-мышечной системы.

Регулируемая высота рабочей поверхности стола должна изменяться в пределах от 680 до 800 мм, а конструкция рабочего стола должна обеспечивать возможность размещения на рабочей поверхности необходимого комплекта оборудования и документов с учетом характера выполняемой работы. Размеры рабочей поверхности стола должны быть: глубина - не менее 600 мм, ширина - не менее 1200 мм. Дисплей на рабочем месте оператора должен располагаться так, чтобы изображение в любой его части было различимо без необходимости поднять или опустить голову. Дисплей на рабочем месте должен быть установлен ниже уровня глаз оператора. Угол наблюдения экрана оператором относительно горизонтальной линии взгляда не должен превышать 60° . Клавиатура должна иметь возможность свободного перемещения. Клавиатуру следует располагать на поверхности стола на расстоянии от 100 до 300 мм от переднего края, обращенного к оператору, или на специальной регулируемой по высоте рабочей поверхности, отделенной от основной столешницы.

Кресло оператора может быть с профилированными и не профилированными элементами согласно ГОСТ 21889-76 «Система "Человек-машина". Кресло человека-оператора» [24]. Поверхность сиденья может быть плоской с наклоном $0-5^\circ$, или профилированной с углами наклона сиденья. Опорная плоскость сиденья может быть плоской или профилированной с радиусом кривизны поясничной опоры, равным 460 мм, радиусом изгиба для грудного отдела позвоночника, равным 620 мм и другими точками изгиба.

При размещении рабочих мест необходимо исключить возможность прямой засветки экрана источником естественного освещения. При выполнении выпускной квалификационной работы правовых и организационных нарушений по указанным требованиям не было выявлено, рабочее место было оборудовано согласно всем нормам и правилам.

3.3 Производственная безопасность

В данном подразделе приведен анализ вредных и опасных факторов, которые воздействуют на разработчиков программного обеспечения, выполняющих работы на своих рабочих местах. Все выявленные факторы приведены в таблице 27 .

Таблица 27 – Возможные опасные и вредные производственные факторы на рабочем месте «офис»

Факторы (ГОСТ 12.0.003-2015)	Нормативные документы
Вредные факторы	
Статические физические перегрузки, связанные с рабочей позой	ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» [37]
Монотонный режим работы	Р 2.2.2006-05 «Руководство по гигиенической оценке факторов рабочей среды и трудового процесса» [33]
Отсутствие или недостаток освещения рабочей зоны	СП 52.13330.2016 «Естественное и искусственное освещение» [34]
Отклонение показателей микроклимата помещения	СанПин 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания» [35]
Опасные факторы	
Опасность поражения электрическим током	ГОСТ Р 12.1.019-2017 «Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты» [36]

3.3.1 Статические физические перегрузки

Работа программиста является малоподвижной деятельностью, а значит может вызывать гиподинамию. Как известно, ограничение физической активности приводит к нарушениям в работе опорно-двигательного аппарата, сердечно-сосудистой системы, желудочно-кишечного тракта. Все эти нарушения оказывают значительное влияние на состояние организма работника и негативно сказываются как на его повседневной жизни, так и на качестве его работы.

Для профилактики гиподинамии следует предпринимать следующие

меры:

- увеличение физической активности;
- регулярные перерывы в работе для небольшой разминки.

3.3.2 Монотонный режим работы

Монотонный труд является частью практически любого процесса разработки, так как каждый специалист выполняет ряд каких-либо действий, связанных с его задачами, что в свою очередь ведет к снижению работоспособности и производительности труда, потери концентрации и быстрому утомлению работников.

Согласно Р 2.2.2006-05 [26], при оптимальном классе труда время пассивного наблюдения, выполнение монотонных действий должно составлять менее 75% от общего времени работы сотрудника.

При постоянном взаимодействии с компьютером (набор текстов, ввод данных и т.п.), при исключении возможности периодического переключения на другие виды трудовой деятельности, не связанные с ПЭВМ, рекомендуется организация перерывов по 10 - 15 мин. через каждые 45 - 60 мин. работы.

3.3.3 Отсутствие или недостаток освещенности рабочей зоны

Недостаточная освещенность рабочей зоны является вредным производственным фактором, приводящим к быстрому утомлению и снижению работоспособности человека на предприятии. При недостаточной освещенности помещения человек быстрее устает, снижается внимание и концентрация. Продолжительная работа в условиях низкой освещенности приводит к ухудшению зрения.

Нормы естественного, искусственного и совместного освещения регламентируются СП 52.13330.2016 «Естественное и искусственное освещение» [27]. Разработка программного обеспечения относится к работам высокой точности (наименьший или эквивалентный объект различения 0,30 – 0,50 мм), разряд Б, подразряд 1, относительная продолжительность зрительной

работы при направлении зрения на рабочую поверхность не менее 70%. В таблице 28 представлены требования к искусственному освещению рабочего помещения.

Таблица 28 – Требования к искусственному освещению рабочего помещения

Искусственное освещение			
Освещенность на рабочей поверхности от системы общего освещения, лк	Цилиндрическая освещенность, лк	Объединенный показатель UGR, не более	Коэффициент пульсации освещенности, Кп, %, не более
400	100	19	15

Искусственное освещение в помещениях для эксплуатации ПК должно осуществляться системой общего равномерного освещения. В случаях преимущественной работы с документами, следует применять системы комбинированного освещения (к общему освещению дополнительно устанавливаются светильники местного освещения).

Яркий свет в зоне периферийного зрения заметно увеличивает напряжение глаз. Для снижения влияния вредного фактора недостаточной освещенности необходимо, чтобы уровень освещения рабочего пространства приблизительно совпадал с яркостью дисплея. Проблему недостаточной освещенности помещения можно решить при помощи установки дополнительных осветительных приборов.

3.3.4 Отклонение показателей микроклимата

Температура воздуха, технологического оборудования, его влажность, скорость движения в помещении, тепловое облучение – это показатели, которые характеризуют микроклимат в офисных помещениях. Причинами отклонения показателей микроклимата может служить поломка кондиционера, его несвоевременная чистка, недостаточное осушение или увлажнение воздуха в помещении. Неподходящий микроклимат в помещении снижает работоспособность, повышает утомляемость и может привести к снижению иммунитета и простудным заболеваниям.

Разработка информационной системы имеет категорию работ Іб и согласно СанПин 1.2.3685-21 [35] в тёплый период года при категории работ Іб диапазон ниже оптимальных значений 20 - 21,9 °С, а диапазон температуры выше оптимальных значений 24.1 – 28 °С, а относительная влажность должна быть 15-75% независимо от времени года, при этом скорость движения воздуха 0.1 – 0.3м/с.

Микроклимат рабочего места должен регулироваться центральным отоплением и приточно-вытяжной вентиляцией. Для улучшения показателей микроклимата необходимо обеспечить правильное кондиционирование и вентиляцию воздуха, отопление помещений. В целях профилактики могут быть использованы средства местного кондиционирования воздуха, помещения для отдыха и обогрева, регламентация времени работы, увлажнения воздуха.

3.3.5 Опасность поражения электрическим током

Работа программиста происходит в непосредственной близости от электрических сетей и приборов, поэтому работник должен с осторожностью обращаться с электропроводкой и компьютером, а также помнить об опасности поражения электрическим током.

Покрытие полов следует делать из однослойного линолеума, что снизит величины зарядов статического электричества. Несмотря на то, что эти величины безопасны для здоровья человека, вычислительная техника подвергается опасности при воздействии зарядов такого рода.

Поражение электрическим током является опасным производственным фактором. Вероятность поражения повышается при:

- повышенной влажности воздуха в помещении (более 75%);
- высокой температуре воздуха и поверхностей (более 35 °С);
- наличии токопроводящей пыли;
- неверной проектировке рабочего места;

- отсутствии защитных конструкций для проводов;
- наличии посторонних предметов на электроприборах.

Мерами защиты от воздействия электрического тока при неисправности изоляции являются защитное заземление, зануление [28] и использование устройств защитного отключения (УЗО).

3.4 Экологическая безопасность

3.4.1 Воздействие на селитебную зону

Большое количество процессов, операций и материалов, используемых при изготовлении электронных средств, являются источниками огромного количества веществ, имеющих неблагоприятное воздействие на человека и биосферу. При изготовлении элементной базы, электронных изделий, при обработке, выращивании полупроводниковых кристаллов, при изготовлении интегральных схем, в процессе гальванического производства утилизация исходных материалов часто происходит с низким коэффициентом использования, огромное количество их идет в отходы, попадая в атмосферу, гидросферу, загрязняя почву. Таким образом, наряду с истощением природных запасов дефицитных материалов происходит загрязнение окружающей среды, что ведет к губительным последствиям для отдельных экосистем и биосферы в целом.

Согласно ГОСТ Р 53692—2009, вышедшее из строя ПЭВМ и сопутствующая оргтехника относится к IV классу опасности и подлежит специальной утилизации.

Первым этапом является утилизация обезвреженных (инертных) отходов. Во время утилизации может быть произведена переработка бракованных или вышедших из употребления видов продукции, изделий, их составных частей и отходов от них путем разборки (разукрупнения), переплавки, использования других технологий с обеспечением рециркуляции (восстановления) органической и неорганической составляющих. Согласно

СанПиН 1.2.3685-21 концентрация стирола, появляющегося при утилизации должна составлять не более 0.02 мг/м³ в помещении, кадмия, содержащегося в полупроводниках – не более 0.5 мг/м³.

Вторым этапом является безопасное размещение отходов I—IV классов опасности на соответствующих полигонах или уничтожение [29].

3.4.2 Воздействие на литосферу

Загрязняющим литосферу веществом является АБС пластик, из которого сделано большинство корпусов компьютерной техники. Пластик при утилизации или при длительном нахождении в почве в результате выброса выделяет полиизобутилен и фторопласты, предельно-допустимая концентрация которых составляет 5 мг/м³, где методом защиты является хорошо проветриваемое помещение и защитный костюм.

3.4.3 Воздействие на гидросферу

Производственные компании загрязняют гидросферу отходами от производства компьютерной техники. Среди них брак, остаточные и смазочные материалы, в которых содержится ртуть, кадмий и свинец. ПДК данных тяжёлых металлов в водных объектах около зданий производства и утилизации составляет 2 мкг/л. Для уменьшения вреда необходимо использовать специальные фильтры на заводах по производству и переработке техники, чтобы избежать попадание загрязняющего вещества в воду

3.4.4 Воздействие на атмосферу

С увеличением спроса на компьютерную технику увеличивается количество компаний-производителей данной продукции, что приводит к разрастанию объёма выбросов углекислого газа в атмосферу. Для уменьшения

выбросов необходимо на каждый завод устанавливать специальные фильтры, с помощью которых воздух будет очищаться от вредных примесей.

Пользователям компьютерной техники для уменьшения углеродного следа рекомендуется использовать энергосберегающие лампы, вовремя выключать свет и компьютерную технику, и если техника неисправна, то следует отнести её в специальное учреждение для последующей переработки или утилизации.

3.5 Безопасность в чрезвычайных ситуациях

Возможные ЧС:

- Наводнение, ураган, землетрясение;
- Эпидемия;
- Пожар.

При работе с вычислительной техникой наиболее типичной чрезвычайной ситуацией является пожар, так как в современных ЭВМ очень высокая плотность размещения элементов электронных схем. К причинам электрического характера можно отнести короткое замыкание, искрение, статическое электричество.

- Для предотвращения возникновения пожара необходимо:
- Регулярно проводить инструктажи по пожарной безопасности.
- Разместить в помещении план эвакуации и плакаты с краткой информацией о действиях, предпринимаемых при возникновении пожара.
- Соблюдать правила и нормы при монтаже электрических приборов и электрической проводки.
- Оборудовать помещение пожарной сигнализацией, а также средствами тушения пожара.

В случае возникновения пожара каждый сотрудник должен:

- незамедлительно сообщить об этом в пожарную охрану;

- принять меры по эвакуации людей, материальных ценностей согласно плану эвакуации;
- отключить электроэнергию, приступить к тушению пожара первичными средствами пожаротушения.

Возможный пожар на рабочем месте относится к классам А1, А2, Е [30]. Первичные средства пожаротушения являются: огнетушители порошковые переносные с порошками типа АВСЕ, огнетушители углекислотные.

3.6 Выводы по разделу

В результате работы по главе «Социальная ответственность» были выявлены основные нормативные акты для обеспечения безопасности жизнедеятельности на рабочем месте, рассмотрены наиболее значимые опасные и вредные факторы, возникающие при проектировании и разработке сервиса для изучения иностранных языков, описано влияние процесса разработки программного обеспечения на окружающую среду и меры, необходимые для уменьшения влияния вредных и опасных факторов на организм человека и для сокращения негативного влияния процесса разработки программного обеспечения на окружающую среду.

Согласно пункту 1.1.13 ПУЭ-7 рабочая зона является помещением без повышенной опасности. Согласно «Правилам по охране труда при эксплуатации электроустановок» [31] персонал должен иметь первую группу по электробезопасности. Работа в офисе относится к категории тяжести труда Ia [32] – работы выполняются при оптимальных условиях внешней производственной среды и при оптимальной величине физической, умственной и нервно-эмоциональной нагрузки. Рабочая зона относится к категории В по взрывопожарной и пожарной опасности.

Все необходимые требования и нормы безопасности были соблюдены, нарушений по организации рабочего процесса не выявлено.

ЗАКЛЮЧЕНИЕ

В ходе выполнения выпускной квалификационной работы была решена задача разработки и реализации программного модуля, способного определять психологический тип личности человека по его текстовым сообщениям. Были изучены и проанализированы существующие теории и методики психологической типологизации, в частности, теории К. Юнга и Майерс-Бриггс (МВТИ) и пяти факторов личности по типологии Big Five. На этапе проектирования были разработаны алгоритм и архитектура программного модуля, основанного на комбинации лингвистического анализа текста и классификации его по 16 психотипам по шкале МВТИ и по 5 факторам личности по шкале Big Five с помощью нейронной сети. На этапе реализации было создано веб-приложение на языке программирования Python с использованием библиотек Flask, TensorFlow и Keras, которое анализирует информацию с социальной сети ВКонтакте по id(screen name) пользователя по типологиям МВТИ и Big Five.

Разработанный программный модуль может быть использован для различных целей, связанных с психологическим консультированием, образованием, маркетингом, рекрутингом и оптимизацией бизнес-процессов. Пользователь может узнать свой тип личности и типы других людей по текстам, а также оценить и подобрать потенциальных клиентов, сотрудников или партнеров по их типу личности.

В дальнейшем планируется развивать приложение, расширять функционал и внедрять в компаниях, желающих оптимизировать свои бизнес-процессы на основе типа личности клиента или сотрудника.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Майерс, И., Бриггс, К. (1980). Gifts Differing: Understanding Personality Type. Nicholas Brealey Publishing.
2. Gjurković, M., & Šnajder, J. (2018). Reddit: A Gold Mine for Personality Prediction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 1292-1297).
3. Michael S. Chmielewski, Theresa A. Morgan , Five-Factor Model of Personality, [Электронный ресурс] URL: https://link.springer.com/referenceworkentry/10.1007/978-1-4419-1005-9_1226, режим доступа:свободный (20.05.2023)
4. Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144.
5. MBTI (Myers-Briggs Personality Type Indicator). [Электронный ресурс]. - режим доступа: URL: <https://www.mbtionline.com/> свободный (20.05.2023)
6. Weis, L. (n.d.). Raters' bias in personality test scoring. Psychological Reports, 60(3), 879-884.
7. Chen, C. L., & Chen, W. (2017). Sentiment analysis: A review of methods and applications. International Journal of Information Management, 37(2), 77-92.
8. Kim, S., & Hovy, E. (2006). Automated identification of English text subjectivity and sentiment. Proceedings of the 2006 conference on empirical methods in natural language processing, 462-469.
9. [TensorFlow](https://www.tensorflow.org/?hl=ru) [Электронный ресурс]. - режим доступа: URL: <https://www.tensorflow.org/?hl=ru> свободный (20.03.2023)
10. [CUDA Toolkit - Free Tools and Training | NVIDIA Developer](#), [Электронный ресурс]. - режим доступа: URL:

- <https://developer.nvidia.com/cuda-toolkit>, свободный (20.03.2023)
11. Keras API reference [Электронный ресурс] / Keras. Режим доступа <https://keras.io/api/>, свободный. (дата обращения: 20.05.2023)
 12. [Flask — Flask Documentation \(2.3.x\) \(palletsprojects.com\)](https://flask.palletsprojects.com/en/2.3.x/) [Электронный ресурс] / Режим доступа: URL: <https://flask.palletsprojects.com/en/2.3.x/> свободный. (дата обращения: 20.03.2023)
 13. Visual Studio, [Электронный ресурс] / Режим доступа: URL: <https://visualstudio.microsoft.com/ru/> свободный. (дата обращения: 20.03.2023)
 14. Anaconda, [Электронный ресурс] / Режим доступа: URL: <https://www.anaconda.com/> свободный. (дата обращения: 20.03.2023)
 15. Tensorflow-GPU: How to Install Tensorflow with NVIDIA CUDA, cuDNN and GPU support on Windows [Электронный ресурс] / Medium. Режим доступа: URL: <https://medium.com/analytics-vidhya/tensorflow-gpu-how-to-install-tensorflow-with-nvidia-cuda-cudnn-and-gpu-support-on-windows-6158cffc1c29/>, свободный. (дата обращения: 20.05.2023)
 16. API VK для разработчиков, [Электронный ресурс] / Режим доступа: URL: <https://dev.vk.com/reference>, свободный. (дата обращения: 20.03.2023)
 17. Stillwell, D. J., & Kosinski, M.. myPersonality Project website, 2015, [Электронный ресурс] – Режим доступа: URL: <https://sites.google.com/michalkosinski.com/mypersonality> свободный. (дата обращения: 20.05.2023)
 18. Zhenkun Zhou, Ke Xu, Jichang Zhao, Extroverts Tweet Differently from Introverts in Weibo, 2017, [Электронный ресурс] URL: https://www.researchgate.net/publication/315456499_Extroverts_Tweet_Differently_from_Introverts_in_Weibo -. Режим доступа: свободный. (дата обращения: 20.05.2023)
 19. Литвинова Т.А., Рыжкова Е.С. Электронный корпус письменных текстов RusNeuroPsych: состав, структура и возможности использования, 2016, [Электронный ресурс] URL: <https://st->

- hum.ru/en/node/444 : -. Режим доступа: свободный. (дата обращения: 20.05.2023)
20. MITCHELL J, (MBTI) Myers-Briggs Personality Type Dataset, [Электронный ресурс] URL: <https://www.kaggle.com/datasets/datasnaek/mbti-type> -. Режим доступа: свободный. (дата обращения: 20.05.2023)
 21. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing, [Электронный ресурс] URL: <https://huggingface.co/microsoft/mdeberta-v3-base> -. Режим доступа: свободный. (дата обращения: 20.05.2023)
 22. Трудовой кодекс Российской Федерации» от 30.12.2001 N 197-ФЗ (ред. от 19.12.2022).
 23. ГОСТ Р 50923-96 «Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200025975> (дата обращения: 22.04.2023).
 24. ГОСТ 21889-76 «Система "Человек-машина". Кресло человека-оператора» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200012832> (дата обращения: 22.04.2023).
 25. ГОСТ 12.0.003-2015 «Система стандартов безопасности труда. Опасные и вредные факторы» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200136071> (дата обращения: 22.04.2023).
 26. МР 2.2.9.2311-07 «Профилактика стрессового состояния работников при различных видах профессиональной деятельности» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200072234> (дата обращения: 22.04.2023).
 27. СП 52.13330.2016 «Естественное и искусственное освещение» [Электронный ресурс]. – Режим доступа: <http://sniprf.ru/sp52-13330-2016> (дата обращения: 22.04.2023).

28. ГОСТ Р 12.1.019-2017 «Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты» [Электронный ресурс]. – Режим доступа: <https://protect.gost.ru/default.aspx/v.aspx?control=7&id=231949> (дата обращения: 22.04.2023).
29. ГОСТ Р 53692-2009 «Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов» [Электронный ресурс]. – Режим доступа: <https://internet-law.ru/gosts/gost/50159/> (дата обращения: 22.04.2023).
30. Федеральный закон от 22.07.2008 N 123-ФЗ (ред. от 30.04.2021) «Технический регламент о требованиях пожарной безопасности» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/902111644> (дата обращения: 22.04.2023).
31. Приказ Минтруда России от 15 декабря 2020 года N 903н «Об утверждении Правил по охране труда при эксплуатации электроустановок» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/573264184> (дата обращения: 22.04.2023).
32. Приказ Минтруда России от 24.01.2014 N 33н (ред. от 27.04.2020) «Об утверждении Методики проведения специальной оценки условий труда, Классификатора вредных и (или) опасных производственных факторов, формы отчета о проведении специальной оценки условий труда и инструкции по ее заполнению» [Электронный ресурс]. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_158398/ (дата обращения: 22.04.2023).
33. Р 2.2.2006-05 ГИГИЕНА ТРУДА. Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. Критерии и классификация условий труда [Электронный ресурс] / Электронный фонд правовых и нормативно-технических документов URL: <https://docs.cntd.ru/document/1200040973> (дата обращения: 22.04.23).

34. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95* (с Изменениями N 1, 2) [Электронный ресурс] / Электронный фонд правовых и нормативно-технических документов URL: <https://docs.cntd.ru/document/456054197> (дата обращения: 22.04.23).
35. СП 51.13330.2011 Защита от шума. Актуализированная редакция СНиП 23-03-2003 (с Изменениями N 1, 2) [Электронный ресурс] / Электронный фонд правовых и нормативно-технических документов URL: <https://docs.cntd.ru/document/1200084097> (дата обращения: 22.04.23).
36. ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление (с Изменением N 1) [Электронный ресурс] / Электронный фонд правовых и нормативно-технических документов URL: <https://docs.cntd.ru/document/5200289> (дата обращения: 22.04.23).
37. ГОСТ 12.0.003-2015 «Система стандартов безопасности труда. Опасные и вредные факторы» [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200136071> (дата обращения: 22.04.2022).

ПРИЛОЖЕНИЕ А

Предобработка данных

```
# function to extract topic from URL
def extract_topic(url):
    try:
        response = requests.get(url)
        html_content = response.text
        match = re.search('<title>(.*?)</title>', html_content)
        if match:
            return match.group(1)
        else:
            return ""
    except Exception:
        last_part = os.path.splitext(url.split("/")[-1])[0]
        return f"{last_part}"

# function to replace URLs in input text
def replace_urls(text):
    urls = re.findall('(P<url>https?://[^\s]+)', text)
    for url in urls:
        topic = extract_topic(url)
        text = text.replace(url, f"url:{topic}")
    return text

tt = nltk.tokenize.TweetTokenizer()

def emoji_and_demoji(text, replace_emoji_with="<EMOJI>"):
    tokens = tt.tokenize(text)
    final_tokens = []
    for t in tokens:
        if t in emoji.EMOJI_DATA :
            if replace_emoji_with:
                final_tokens.append(replace_emoji_with)
            else:
                final_tokens.append(t)
            final_tokens.append(emoji.demojize(t).strip(':'))
        else:
            final_tokens.append(t)
    return " ".join(final_tokens)
```

ПРИЛОЖЕНИЕ Б

Сборка модели

```
from transformers import TFDebertaV2Model

def build_model(max_len=512):
    input_word_ids = tf.keras.layers.Input(shape=(max_len,), dtype=tf.int32,
name='input_word_ids')
    input_mask = tf.keras.layers.Input(shape=(max_len,), dtype=tf.int32, name='input_mask')
    input_type_ids = tf.keras.layers.Input(shape=(max_len,), dtype=tf.int32, name='input_type_ids')

    bert_layer = TFDebertaV2Model.from_pretrained("microsoft/mdeberta-v3-base")

    sequence_output = bert_layer({'input_ids': input_word_ids, 'attention_mask': input_mask,
'token_type_ids': input_type_ids})
    clf_output = sequence_output.last_hidden_state[:, 0, :]

    out = tf.keras.layers.Dense(len(np.unique(train_labels)), activation='sigmoid')(clf_output)

    model = tf.keras.Model(inputs=[input_word_ids, input_mask, input_type_ids], outputs=out)
    model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=2e-7),
loss='binary_crossentropy',
metrics=["accuracy",tf.keras.metrics.AUC(),tf.keras.metrics.Precision(),tf.keras.metrics.Recall()
,tf.keras.metrics.TruePositives(thresholds=None, name=None, dtype=None),
tf.keras.metrics.TrueNegatives(thresholds=None, name=None, dtype=None),
tf.keras.metrics.FalsePositives(thresholds=None, name=None, dtype=None),
tf.keras.metrics.FalseNegatives(thresholds=None, name=None, dtype=None)])

    return model

model = build_model(max_len=max_len)
```

ПРИЛОЖЕНИЕ В

Пользовательский callback на keras

```
from keras.callbacks import Callback
class MultipleValidationCallback(Callback):
    def __init__(self, validation_data, n, model_file, patience ):
        super().__init__()
        self.validation_data = validation_data
        self.n = n # number of times to validate per epoch
        self.model_file = model_file
        self.best_val_loss = float('inf')
        self.patience = patience # added patience attribute

    def on_epoch_end(self, patience):
        for i in range(self.n):
            # validate the model on the validation data
            validation_results = self.model.evaluate(self.validation_data[0],
            self.validation_data[1],
            verbose=0)
            # print the validation metrics
            print(f'Validation results (iteration {i+1}/{self.n}): {validation_results}')

            # save the model if the validation loss has improved
            if validation_results[0] < self.best_val_loss:
                self.best_val_loss = validation_results[0]
                self.model.save(self.model_file)
                self.patience = patience # reset patience when a better model is saved
            else:
                self.patience -= 1
                if self.patience == 0:
                    print(f'Validation loss has not improved for {self.patience} iterations. Stopping.')
                    self.model.stop_training = True
```