

Школа – Инженерная школа информационных технологий и робототехники
 Направление подготовки – 09.03.02 «Информационные системы и технологии»
 Отделение школы (НОЦ) – Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Система локальной проверки оригинальности студенческих работ

УДК 004..732:004.083.7:378.146.261

Студент

Группа	ФИО	Подпись	Дата
8И8А	Мукомберо Хоуп		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко Ирина Валериевна	К.Т.Н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН УОД	Рыжакина Татьяна Гавриловна	К.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	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко Ирина Валериевна	К.Т.Н.		

Томск – 2022г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

Код компетенции	Наименование 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
УК(У)-7	Способен поддерживать должный уровень физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности
УК(У)-8	Способен создавать и поддерживать в повседневной жизни и в профессиональной деятельности безопасные условия жизнедеятельности для сохранения природной среды, обеспечения устойчивого развития общества, в том числе при угрозе и возникновении чрезвычайных ситуаций и военных конфликтов
УК(У)-9	Способен проявлять предприимчивость в практической деятельности, в т.ч. в рамках разработки коммерчески перспективного продукта на основе научно-технической идеи
УК(У)-10	Способен принимать обоснованные экономические решения в различных областях жизнедеятельности
УК(У)-11	Способен формировать нетерпимое отношение к коррупционному поведению
ОПК(У)-1	Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
ОПК(У)-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их 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 технической документации, связанной с профессиональной деятельностью с использованием стандартов, норм и правил

Код компетенции	Наименование компетенции
ОПК(У)-5	Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать алгоритмы и программы, пригодные для практического применения в области информационных систем и технологий
ОПК(У)-7	Способен осуществлять выбор платформ и инструментальных программно-аппаратных средств для реализации информационных систем
ОПК(У)-8	Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем
ПК(У)-1	Способен выполнять интеграцию программных модулей и компонент
ПК(У)-2	Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению информационных систем
ПК(У)-3	Способен создавать техническую документацию на продукцию в сфере информационных технологий, управлять технической информацией
ПК(У)-4	Способен выполнять работы по обеспечению функционирования баз данных и обеспечению их информационной безопасности
ПК(У)-5	Способен проводить, оценивать и следить за выполнением концептуального, функционального и логического проектирования систем малого и среднего масштаба и сложности

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: 09.03.02 Информационные системы и технологии
 Отделение школы (НОЦ): Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП

_____ _____ Цапко И. В.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы

Студенту:

Группа	ФИО
8И8А	Мукомберо Хоуп

Тема работы:

Система локальной проверки оригинальности студенческих работ	
Утверждена приказом директора (дата, номер)	№34–61/с от 03.02.2022 г.

Срок сдачи студентом выполненной работы:	06.06.2022 г.
--	---------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Система предназначена для выявления подобных работ студентов, расположенных на локальном диске, с указанием степени их оригинальности и последующим выделением совпадающих работ в отдельные кластеры.</p>
---	---

<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> 1. Исследование теоретические основы плагиата 2. Анализ существующих алгоритмов, библиотек и систем проверки на заимствование 3. Реализация системы проверки оригинальности работ 4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение 5. Социальная ответственность
<p>Перечень графического материала <i>(с точным указанием обязательных чертежей)</i></p>	<p>Презентация в формате *.pptx</p>
<p>Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i></p>	
<p>Раздел</p>	<p>Консультант</p>
<p>Финансовый менеджмент</p>	<p>Рыжакина Татьяна Гавриловна</p>
<p>Социальная ответственность</p>	<p>Мезенцева Ирина Леонидовна</p>
<p>Названия разделов, которые должны быть написаны на русском и иностранном языках:</p>	
<p>Все разделы должны быть написаны на русском языке.</p>	

<p>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</p>	<p>24.01.2022 г.</p>
--	----------------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
<p>Доцент ОИТ ИШИТР</p>	<p>Цапко И. В.</p>	<p>к.т.н.</p>		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
<p>8И8А</p>	<p>Мукомберо Хоуп</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 – 09.03.02 Информационные системы и технологии
 Уровень образования – Бакалавриат
 Отделение школы (НОЦ) – Отделение информационных технологий
 Период выполнения – Весенний семестр 2021/2022 учебного года

Форма представления работы:

Бакалаврская работа

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

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	06.06.2022 г.
--	---------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
30.05.2022	Основная часть	75
28.05.2022	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
28.05.2022	Социальная ответственность	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко И. В.	К.Т.Н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко И. В.	К.Т.Н.		

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

Студенту:

Группа	ФИО
8И8А	Мукомберо Хоуп

Школа	Инженерная школа информационных технологий и робототехники	Отделение школы (НОЦ)	Отделение информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Оклад руководителя (к.т.н., доцент) – 33664 руб. Оклад студента – 21760 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	Премиальный коэффициент 30%; Коэффициент доплат и надбавок 40%; 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 30%

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	Проведение предпроектного анализа. Определение целевого рынка и проведение его сегментирования. Выполнение SWOT-анализа проекта
2. <i>Планирование и формирование бюджета научных исследований</i>	Формирование плана и графика работы: - определение структуры работ; - определение трудоёмкости; - построение графика Ганта. Формирование бюджета затрат: - материальные затраты; - затраты на спец.оборудование; - заработная плата; - отчисления во внебюджетные фонды; - накладные расходы.
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	Определение и оценка показателя ресурсоэффективности исследования

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Оценка конкурентоспособности технических решений</i>
2. <i>Матрица SWOT</i>
3. <i>Диаграмма Ганта</i>

- | |
|--|
| 4. График проведения и бюджет НИ |
| 5. Оценка ресурсной, финансовой и экономической эффективности НИ |

Дата выдачи задания для раздела по линейному графику	03.02.2022
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Рыжакина Татьяна Гавриловна	к.э.н.		03.02.2022

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И8А	Мукомберо Хоуп		03.02.2022

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа		ФИО	
8И8А		Мукомберо Хоуп	
Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

Тема ВКР:

Система локальной проверки оригинальности студенческих работ

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 эксплуатации 	<p><i>Объект исследования:</i> система локальной проверки оригинальности студенческих работ с использованием алгоритма шинглов.</p> <p><i>Область применения:</i> высшие учебные заведения</p> <p><i>Рабочая зона:</i> место оператора ПЭВМ</p> <p><i>Размеры помещения:</i> 4*6 м.</p> <p><i>Количество и наименование оборудования рабочей зоны:</i> 1 ПЭВМ.</p> <p><i>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне:</i> загрузка файлов для проверки схожести документов и контроль параметров процесса кластеризации.</p>
--	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Правовые и организационные вопросы обеспечения безопасности при эксплуатации:</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>Нормы трудового законодательства и организационных мероприятий при работе с ПЭВМ:</p> <ul style="list-style-type: none"> - «Трудовой кодекс Российской Федерации» от 30.12.2001 N 197-ФЗ (ред. от 01.03.2022). - ГОСТ 21889–76 Система «Человек-машина». Кресло человека-оператора. - ГОСТ 12.2.032–78 Рабочее место при выполнении работ сидя. - ГОСТ 12.0.003–2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
<p>2. Производственная безопасность при эксплуатации:</p> <ul style="list-style-type: none"> – Анализ выявленных вредных и опасных производственных факторов – Обоснование мероприятий по снижению воздействия 	<p>Вредные факторы:</p> <ol style="list-style-type: none"> 1. отсутствие или недостатки необходимого искусственного освещения; 2. монотонность труда, вызывающая монотонию; 3. длительность сосредоточенного наблюдения. <p>Опасные факторы:</p>

	<p>1. опасные и вредные производственные факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий.</p> <p>Требуемые средства коллективной и индивидуальной защиты от выявленных факторов: источники света, осветительные приборы, световые проемы, устройства защитного заземления и зануления.</p> <p>Расчет: расчет системы искусственного освещения</p>
3. Экологическая безопасность при эксплуатации	<p>Воздействие на литосферу: утилизация отходов электрооборудования.</p> <p>Воздействие на гидросферу: не выявлено.</p> <p>Воздействие на атмосферу: не выявлено.</p> <p>Воздействие на селитебную зону: не выявлено.</p>
4. Безопасность в чрезвычайных ситуациях при эксплуатации	<p>Возможные ЧС: пожары, обрушение зданий, землетрясения.</p> <p>Наиболее типичная ЧС: пожар.</p>
Дата выдачи задания для раздела по линейному графику	
	15.02.2022

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Мезенцева Ирина Леонидовна			15.02.2022

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И8А	Мукомберо Хоуп		15.02.2022

РЕФЕРАТ

Выпускная квалификационная работа содержит: 87 страниц, 16 рисунка, 24 таблицу, 24 источников, 1 приложение.

Ключевые слова: антиплагиат, алгоритм шинглов, оригинальность, кластеризация, информационные системы.

Актуальность работы: заключается в том, что она решает проблему плагиата неопубликованных студенческих работ (отчетов), представляемых преподавателям внутри вуза и, следовательно, повышает качество работы, выполняемой студентами в образовательных учреждениях.

Объектом исследования выпускной квалификационной работы является система предназначена для выявления подобных работ студентов, расположенных на локальном диске, с указанием степени их оригинальности и последующим выделением совпадающих работ в отдельные кластеры.

Цель работы: разработка программы локальной проверки оригинальности студенческой работы.

В процессе выполнения работы была реализована система, позволяющая сравнивать файлы студенческих работ, расположенные на локальном диске, и определять подобные из них указанием процента заимствования. В систему был успешно добавлен способ кластеризации результатов и включение одновременного много файлового сравнения. Функция фильтрации результатов на основе процентного сходства сработала эффективно.

Среда разработки: VS Code.

Область применения: высшие учебные заведения.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ССБТ – система стандартов безопасности труда;

ГОСТ – межгосударственный стандарт;

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

ПО – программное обеспечение;

ИС – информационная система;

GUI – Graphical User Interface, графический интерфейс пользователя;

SDK – Software Development Kit;

UML – Unified Modelling Language;

ПК – персональный компьютер;

ЦП – центральный процессор;

ПЭВМ – персональная электро-вычислительная машина;

ЧС – чрезвычайная ситуация.

ОГЛАВЛЕНИЕ

РЕФЕРАТ	11
ВВЕДЕНИЕ	16
1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПЛАГИАТА	18
1.1 Основные понятия и определения	18
1.2 Методы обнаружения плагиата	19
1.3 Анализ существующих алгоритмов	20
1.4 Анализ существующих библиотек	24
1.5 Анализ существующих систем проверки на заимствование	27
2. РЕАЛИЗАЦИЯ СИСТЕМЫ ПРОВЕРКИ ОРИГИНАЛЬНОСТИ РАБОТ	34
2.1 Пользовательские требования к разрабатываемой системе	34
2.2 Варианты использования системы	34
2.3 Используемые технологии	35
2.4 Алгоритм работы программы	36
2.5 Алгоритм шинглов	39
3. РЕЗУЛЬТАТЫ РАБОТЫ СИСТЕМЫ	42
4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	46
4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения	46
4.1.1 Потенциальные потребители результатов исследования	46
4.2 Анализ конкурентных технических решений	48
4.3 SWOT– анализ	49
4.4 Планирование работ по научно-техническому исследованию	53

4.4.1 Структура работ в рамках научного исследования	53
4.4.2 Определение трудоемкости выполнения работ	54
4.4.3 Разработка графика проведения научного исследования	55
4.5 Бюджет научно-технического исследования (НТИ)	61
4.5.1 Расчет материальных затрат НТИ	61
4.5.2 Расчет затрат на специальное оборудование для научных работ	62
4.5.3 Основная заработная плата исполнителя темы	62
4.5.4 Расчет дополнительной заработной платы	65
4.5.5 Отчисления во внебюджетные фонды	65
4.5.6 Накладные расходы	66
4.5.7 Формирование бюджета затрат научно-исследовательского проекта	67
4.6 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	68
4.7 Вывод на раздел	71
5. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	72
5.1 Введение	72
5.2 Правовые и организационные вопросы обеспечения безопасности	73
5.3 Производственная безопасность	75
5.4 Анализ выявленных вредных и опасных факторов	76
5.4.1 Отсутствие или недостаток необходимого искусственного освещения	76
5.4.2 Монотонность труда, вызывающая монотонию	77
5.4.3 Длительность сосредоточенного наблюдения	77
5.5 Экологическая безопасность	79
5.6 Безопасность в чрезвычайных ситуациях	80
5.7 Выводы по разделу	81

ЗАКЛЮЧЕНИЕ	83
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	84
Приложение А	87

ВВЕДЕНИЕ

Во всех университетах мира преподаватели раздают домашние задания, лабораторные практикумы и различные формы курсовой работы, которые студенты должны выполнить в рамках семестровой оценки. Цель состоит в том, чтобы учащиеся индивидуально выполняли эти задания и, следовательно, учились исследовать и решать проблемы. Однако из-за заранее разработанных курсов, которые имеют одни и те же задания в течение нескольких лет, ученики младших курсов склонны копировать эти задания у старших учеников и своих товарищей, которые честно выполнили задания. Преподаватели оказываются с огромным количеством плагиатных заданий, но без быстрого и эффективного способа отсеять списывания от честных студентов. Выпускная квалификационная работа посвящена разработке офлайн-системы проверки оригинальности студенческой работы. Система позволит осуществлять процесс проверки подлинности неопубликованных студенческих работ (отчетов), представляемых преподавателям внутри вуза. При сдаче заданий студентами возникает необходимость проверки на уникальность работ студентов, чтобы свести к минимуму копирование. В Интернете можно найти программное обеспечение для защиты от плагиата, которое проверяет материалы на оригинальность в опубликованных материалах. Иногда в этом нет необходимости для локализованных заданий, выдаваемых учащимся, которые вместо этого требуют проверки на уникальность по сравнению с их автономными аналогами. Этим системам не хватает функционала для проверки оригинальности работы неопубликованных материалов на уроках, когда учащиеся представляют их преподавателям для оценивания. Это создало культуру повторного использования старых отчетов старшеклассников для выполнения заданий. Негативные последствия этой привычки проявляются, когда студенты испытывают затруднения при написании курсовых и выпускных квалификационных работ, требующих проверки на плагиат, так как они, так и не овладели навыками исследовательской и написания оригинальной работы.

Необходимость проверки уникальности содержания документов является центральной проблемой во многих профессиональных кругах. В академических учреждениях это позволяет поддерживать высокие стандарты исследовательской работы и заданий, представляемых студентами. Улучшение выявления плагиата повысит и облегчит обучение, что находит отражение в качестве учащихся в результате образовательного процесса [2].

Необходимо разработать систему, предназначенную для выявления похожих работ учащихся, находящихся на локальном диске, с указанием степени их оригинальности и последующего разделения совпадающих работ на отдельные кластеры. Программа должна уметь выбирать файл в формате *.docx и сравнивать его с другим, выдавая итоговое сходство в процентах.

Целью данной работы является разработка программы локальной проверки оригинальности студенческой работы.

Актуальность данной работы заключается в том, что она решает данную проблему плагиата и, следовательно, повышает качество работы, выполняемой студентами в образовательных учреждениях. Минимальный одноранговый плагиат поощряет дальнейшее исследование и использование уникальных идей для решения проблем, тем самым достигая целей образования.

1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПЛАГИАТА

1.1 Основные понятия и определения

Плагиат – это когда вы используете чужие слова или идеи без указания источника и выдаете их за свои. Согласно [3], существует некоторые типы плагиата, с которым можно столкнуться, перечисленные ниже.

Дословный плагиат, также известный как плагиат копирования и вставки, включает в себя прямое копирование и вставку текста из источника без указания автора. Чтобы законно использовать точные слова автора, необходимо процитировать первоисточник, поместив первоисточник в кавычки и включив цитату в текст. Создание совершенно нового текста путем копирования фраз и понятий из нескольких источников называется лоскутным или мозаичным плагиатом. Прилагаемые усилия по перефразированию оправдывают простое избегание этого с самого начала. Перефразирование плагиата является одним из наиболее популярным типом плагиата. Ошибочное предположение состоит в том, что можно переписывать идеи своими словами, поэтому это оправдывает превращение идеи в свою. Кредит и цитаты всегда должны быть предоставлены первоисточнику. Глобальный плагиат – это когда человек полностью берет чужую работу и использует ее как свою. Наиболее частые формы его проявления – это когда человек находит текст в сети и выдает его как свою работу, а также если просит коллегу написать за него ваше эссе и выдать его за свое. Вопиющий преднамеренный обман в глобальном плагиате делает его одним из самых серьезных видов плагиата. Повторное использование ранее представленной работы или повторное использование идей, разработанных на основе предыдущих заданий, называется самоплагиатом. Независимо от того, что эта работа принадлежит человеку, повторная отправка ее как нового материала по-прежнему считается академической нечестностью, потому что вы уже получили признание за эту работу. Наша работа, как и в задании, будет в основном

включать проверку на мозаичный и глобальный плагиат/сходство файлов в выбранной папке, таких текстовых документов, как студенческие отчеты.

1.2 Методы обнаружения плагиата

Компьютерные методы обнаружения сходства делятся на три группы, а именно: локальное сходство, сходство на основе текста и глобальное сходство (Рисунок 1). Локальные сходства требуют сканирования локальных разделов или сегментов на предмет сходства. Обнаружение глобального сходства включает в себя попытку сопоставить весь документ с другим с точки зрения стиля написания, который может быть таким же, или плагиатом цитирования. Однако между локальным и глобальным методами существует анализ вхождений терминов, который, по сути, пытается найти сопоставление больших фраз и подстрок. Типы разделов сопоставляются с другими в другом документе, так что подстрока/набор слов/некоторые важные наборы слов являются ключевыми словами из одного документа, совпадающими с другими в другом документе. Самый высокий уровень обнаружения достигается при совпадении одной и той же строки, что в основном наблюдается при копировании и вставке части текста. Отпечатки пальцев требуют выбора нескольких подстрок из пакета документов, которые затем сравниваются с отпечатками пальцев для всех документов в пакете. Совпадения отражают сходство документов. Деревья/массивы суффиксов обычно используются для обнаружения подобия [16].

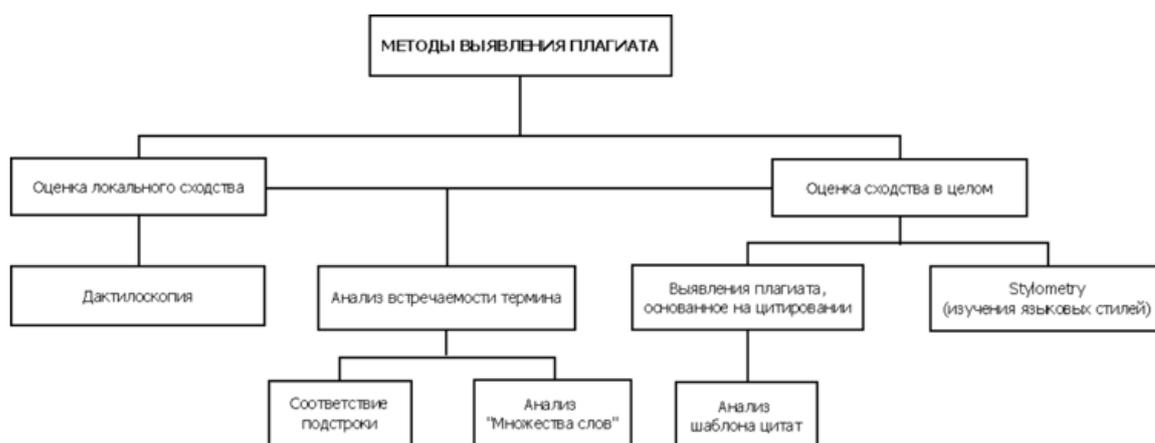


Рисунок 1 — Методы выявления плагиата

Обычно используемый подход для сопоставления похожих документов основан на подсчете максимального количества общих слов между документами.

1.3 Анализ существующих алгоритмов

За прошедшие годы было разработано несколько алгоритмов для анализа сходства текстов. Методы схожести строк обозначаются в согласовании с качествами операций.

Редактировать на базе расстояния: это методы, которые планируют численность операций, важных для переустройства одной строчки в иную. Большая численность операций предполагает наименьшее однообразие между двумя строчками. В данных методах любому знаку индекса строчки придается однообразный смысл. Примерами алгоритмов данной категории считаются: расстояние Хэмминга, расстояние Левенштейна, расстояние Яро-Винклера и традиционное косинусное однообразие.

На базе токенов: в эти способы вводится комплект токенов взамен полных строк. Концепция заключается в том, чтобы находить аналогичные токены в любом комплекте. Чем больше численность совокупных токенов, тем больше однообразие меж наборами. Строчка преобразуется в наборы методом деления с внедрением разделителя. Этим образом, мы можем конвертировать предложение в токены текстов или же знаки n -грамм. Обратите внимание, токены различной длины имеют однообразное смысл. Примеры алгоритмов данной категории: Индекс Жаккарда (Jaccard Index) и Соренсен-Дайс (Sorensen-Dice).

На базе очередности: тут однообразие может быть связано с совместными подстроками между двумя строчками. Различные алгоритмы пытаются отыскать современную расширенную очередность, находящуюся в обеих строчках, чем больше этих последовательностей найдено, тем выше показатель однообразия. Следует обратить внимание, что композиции знаков одинаковой длины имеют схожий смысл.

1.3.1 Классическое косинусное сходство

Классическое косинусное сходство — это мера подобия, которая измеряет косинус угла между двумя векторами, спроецированными на

многомерную плоскость [4]. Это суждение основано скорее на ориентации, чем на величине. Для двух векторов атрибутов, A и B , косинусное сходство представлено с использованием скалярного произведения и величины как:

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A||B|} \quad (1)$$

где θ – угол между векторами A и B ;

A и B – двух векторов атрибутов.

При сравнении текстового сходства каждому слову присваивается отдельная координата, а документ представляется вектором количества вхождений каждого слова в документе. Затем косинусное сходство дает полезную меру того, насколько похожими могут быть два документа с точки зрения их предмета и независимо от длины документов [5]. Два пропорциональных вектора имеют косинусное подобие единицу, два ортогональных вектора имеют подобие нулю, а два противоположных вектора имеют подобие минус единица. Косинусное подобие особенно используется в положительном пространстве, где результат четко ограничен в стиле отображения от нуля до единицы. Обычно используемый подход для сопоставления похожих документов основан на подсчете максимального количества общих слов между документами. Но у этого подхода есть существенный недостаток. По мере увеличения размера документа количество общих слов увеличивается, даже если в документах говорится о разных темах. Косинусное подобие помогает преодолеть этот фундаментальный недостаток подхода «подсчет общих слов» или евклидова расстояния.

1.3.2 Коэффициент Жаккарда

Коэффициент Жаккарда, также известный как коэффициент подобия Жаккарда, представляет собой статистику, используемую для оценки сходства и разнообразия выборочных наборов (в нашем случае текстов). Коэффициент Жаккарда — это самый простой и стандартный алгоритм проверки подобия с простой формулой без какой-либо векторизации. Алгоритм Жаккарда не использует косинусное сходство, а вместо этого размечает слова и вычисляет

пересечение по объединению. Для предварительной обработки текста выполняются следующие шаги:

- весь текст в нижнем регистре;
- токенизировать;
- удалить стоп-слова;
- удалить знаки препинания;
- лемматизировать;
- вычислить пересечение/объединение в двух документах.

Для двух документов A и B коэффициент Жаккарда рассчитывается по формуле 2, а визуальное представление переменных показано на рисунке 2.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

где J = Коэффициент Жаккарда;

A = множество 1;

B = множество 2.

Два набора A и B в нашем случае будут представлять два сравниваемых документа. Слова, найденные в обоих документах, будут представлены пересечением, а общее количество уникальных слов будет объединением, как показано на рисунке 2 ниже.

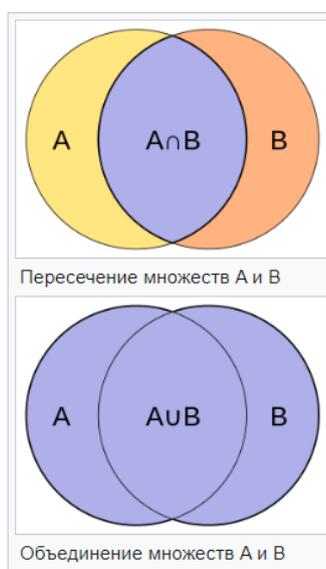


Рисунок 2 — Визуализация переменных, необходимых для расчета индекса Жаккарда.

1.3.3 Мешок слов

Для простого изображения при извлечении информации используется модель мешка слов. Документ представлен в виде набора слов, лишенных грамматического смысла и порядка слов, но сохраняющих свою множественность. Эта модель широко используется при классификации документов. Частота встречаемости слов используется в качестве признака для обучения классификатора. Мешок слов — это представление текста, описывающее вхождение слов в документ. Это включает в себя две вещи:

- словарь известных слов;
- мера наличия известных слов.

Его называют «мешком» слов, потому что любая информация о порядке или структуре слов в документе отбрасывается. Модель касается только того, встречаются ли известные слова в документе, а не где в документе. Интуиция подсказывает, что документы похожи, если они имеют сходное содержание [9]. Далее, что из одного только содержания мы можем кое-что узнать о значении документа. Пакет слов может быть простым или сложным, как вам нравится. Сложность возникает как при принятии решения о том, как спроектировать словарь известных слов (или токенов), так и о том, как оценивать наличие известных слов. Шаги модели мешка слов:

Шаг 1: соберите данные;

Шаг 2: создайте словарный запас;

Шаг 3: создайте векторы документа;

Шаг 4: подсчет слов.

Цель состоит в том, чтобы превратить каждый документ со свободным текстом в вектор, который мы можем использовать в качестве входных или выходных данных для модели машинного обучения. Длина вектора документа равна количеству известных слов. Длина вектора может составлять тысячи или миллионы позиций, если каждый документ содержит очень мало известных слов в словаре, что приводит к вектору с большим количеством нулевых оценок,

называется разреженным вектором или разреженным представлением. Разреженные векторы требуют больше памяти и вычислительных ресурсов при моделировании, а огромное количество позиций или измерений может сделать процесс моделирования очень сложным для традиционных алгоритмов. Таким образом, существует необходимость уменьшить размер словарного запаса при использовании модели мешка слов. В качестве первого шага используются простые методы очистки текста, такие как:

- игнорирование регистра;
- игнорирование знаков препинания;
- игнорирование частых слов, не содержащих много информации,

называемых стоп-словами, таких как «а», «из» и т. д.

- исправление слов с ошибками;
- сокращение слов до их корня с использованием алгоритмов

определения основы.

1.4 Анализ существующих библиотек

Существует несколько библиотек для разных языков, для облегчения сравнения текстов с использованием обширных алгоритмов, доступных пользователям. TextDistance 4.2.2 — библиотека Python для сравнения расстояния между двумя или более последовательностями с помощью многих алгоритмов. Функции библиотеки:

- 30+ алгоритмов;
- Чистая реализация Python;
- Простое использование;
- Сравнение более двух последовательностей;
- Некоторые алгоритмы имеют более одной реализации в одном классе;
- Необязательное использование numru для максимальной скорости.

Алгоритмы в этой библиотеке классифицируются по следующим подгруппам:

- Редактировать на основе;
- На основе токенов;
- Основанный на последовательности;
- На основе сжатия;
- Фонетическая основа .

Token based

Algorithm	Class	Functions
Jaccard index	Jaccard	jaccard
Sørensen–Dice coefficient	Sorensen	sorensen, sorensen_dice, dice
Tversky index	Tversky	tversky
Overlap coefficient	Overlap	overlap
Tanimoto distance	Tanimoto	tanimoto
Cosine similarity	Cosine	cosine
Monge-Elkan	MongeElkan	monge_elkan
Bag distance	Bag	bag

Рисунок 3 — Алгоритмы на основе токенов в библиотеке Python TextDistance

В dart пакет SDK string-similarity 2.0.0 предлагает API-интерфейсы, такие как string-similarity, который находит степень сходства между двумя строками на основе коэффициента Дайса, который в большинстве случаев лучше, чем расстояние Левенштейна. NuGet диспетчер пакетов, позволяющий разработчикам совместно использовать повторно используемый код, также предлагает несколько библиотек, помогающих в разработке методов обнаружения подобия, как показано на рисунке 4.

	StringSimilarity.NET	A .NET port of java-string-similarity.	Score: 3.3 votes (0) 9/14/2018 v 5.0.0
	StringSimilarity	This class library calculates a score from 0 to 1 based on the similarity of two supplied strings.	Score: 2.8 votes (0) 11/15/2020 v 1.0.2.1
	StringSimilarity.NET	A library implementing different string similarity and distance measures. A dozen of algorithms (including Levenshtein edit distance and siblings, Jaro-Winkler, Longest Common Subsequence, cosine similarity etc.) are currently implemented. Based upon F23.StringSimilarity	Score: 2 votes (0) 8/21/2018 v 3.0.0
	Indigo.Net	Universal cheminformatics toolkit	Score: 1.8 votes (0) 12/29/2020 v 1.7.0-beta
	Percolator Matching	A simple dll that contains a matching class to match strings and to calculate the score of similarity between the two strings using the Ratcliff-Obershelp algorithm.	Score: 1.6 votes (0) 5/1/2015 v 1.1.0

Рисунок 4 — Пакеты подобия NuGet

1.5 Анализ существующих систем проверки на заимствование

В процессе выполнения ВКР был проведен анализ существующих систем проверки на заимствование. Были проанализированы как Российские, так и иностранные системы. Результаты сравнения представлены в таблице 1.

Таблица 1 – Сравнение существующих систем антиплагиата

Название системы	Онлайн/Офлайн	Используемый алгоритм	Преимущества	Недостатки	Планы и цены
1. Dupli Checker	Онлайн	Технология на основе искусственного интеллекта	<ul style="list-style-type: none">• Бесплатно• Простота использования• 2 способа проверки на плагиат• 50 проверок на плагиат в день для зарегистрированных пользователей• Несколько форматов файлов (.tex, .txt, .doc, .docx, .odt, .pdf, and .rtf)	<ul style="list-style-type: none">• 1 поиск в день для незарегистрированных пользователей• недоступен в автономном режиме• только проверяет на плагиат опубликованные материалы в Интернете	<ul style="list-style-type: none">• Нет платной версии

2. Copyleaks	Онлайн	Технология на основе искусственного интеллекта	<ul style="list-style-type: none"> • Для образования и бизнеса • Несколько форматов файлов и языков • Разнообразие инструментов 	<ul style="list-style-type: none"> • Вы можете использовать его только после регистрации • Ограничение на количество бесплатных страниц (первые 10) • невозможно сравнить более 2 файлов одновременно на предмет сходства 	<ul style="list-style-type: none"> • Для предприятий в месяц \$566.66. • Для бизнеса начиная с \$ 10,99 для проверки до 25 000 слов. Предоплаченный вариант для бизнеса также доступен здесь.
--------------	--------	--	--	--	---

3. Антиплагиат	Онлайн	Технология на основе искусственного интеллекта и шаг шингла 4	<ul style="list-style-type: none"> • Размер файла до 20 Мб. • Дополнительные параметры типов файлов. • Подсчет цитируемости и само цитирования. 	<ul style="list-style-type: none"> • В течение 6 минут можно проверить не более 1 документа. • Могут обрабатываться ограниченные типы файлов (pdf и txt) и размеры файлов. • Нет доступа к полному отчету о результатах • Не более 3-х источников • Размер файла не более 15Мб. 	<ul style="list-style-type: none"> • Полный тариф на проверку одного документа составляет 472 рубля, срок действия услуги 180 дней.
4. Plagiarism CheckerX	онлайн и офлайн	Технология на основе искусственного интеллекта	<ul style="list-style-type: none"> • предлагает широкие возможности • предлагает автономное параллельное сравнение документов • предлагает автономное массовое сравнение документов 	<ul style="list-style-type: none"> • вам необходимо зарегистрироваться, чтобы использовать программу • не более 5 файлов разрешено для массового сравнения в бесплатной версии 	<ul style="list-style-type: none"> • Бесплатно. • Профессионально \$39.95 • Бизнес \$147.45

5. PaperRater	Онлайн	Grendel алгоритмы машинного обучения и обработки естественного языка	<ul style="list-style-type: none"> • 3 инструмента в 1: корректор и проверка грамматики, построитель словарного запаса, проверка на плагиат • Разработано отраслевыми экспертами • Быстрые результаты 	<ul style="list-style-type: none"> • Нет возможности сохранять отчеты • содержит рекламу • всего 20 страниц на подачу 	<ul style="list-style-type: none"> • Basic - бесплатно, отправка до 5 страниц • Премиум \$7,95 в месяц или \$95,40 в год.
6. Plagiarisma	Онлайн	Технология на основе искусственного интеллекта	<ul style="list-style-type: none"> • Поддерживается более 190 языков • Проверка на плагиат по URL-адресу, в Интернете или при загрузке файла. • Расширения для Firefox и Chrome 	<ul style="list-style-type: none"> • Ограниченные проверки на плагиат в день • требует авторизации • сравнивает только опубликованные материалы 	<ul style="list-style-type: none"> • \$5 за один день до 100 поисков. • Для неограниченных поисков есть подписка за 25 или 30 долларов на 3 или 6 месяцев соответственно. • Для 25 пользователей существует фиксированная ежемесячная

					подписка в размере \$25 для неограниченного количества поисков.
7. Plagiarism Checker	Онлайн		<ul style="list-style-type: none"> • Подробные инструкции • Проверяет не спланировали другие пользователи ваш онлайн-контент 	<ul style="list-style-type: none"> • Полностью онлайн • Поддерживает только браузеры Google или Yahoo 	<ul style="list-style-type: none"> • Бесплатно.
8. Plagium	Онлайн	алгоритм глубокого поиска	<ul style="list-style-type: none"> • Простота в использовании • Имеет 2 типа поиска, быстрый поиск и глубокий поиск. • Бесплатно до 5000 символов каждый раз 	<ul style="list-style-type: none"> • Ограниченные бесплатные функции • проверяет только опубликованный материал только онлайн 	<ul style="list-style-type: none"> • За \$9,99 в месяц вы можете выполнить 287 быстрых и 143 глубоких поиска. • За \$24,99 в месяц вы можете выполнить 699 быстрых и 349 глубоких поисков.

					<ul style="list-style-type: none"> • Подписка за \$99,99 в месяц, которая предлагает 2949 быстрых и 1474 глубоких поиска.
9. PlagScan	Онлайн		<ul style="list-style-type: none"> • Нет подписки для частных пользователей • Возможности интеграции • 3 способа проверки на плагиат 	<ul style="list-style-type: none"> • Полностью онлайн • Относительно сложный интерфейс • Проверяет опубликованный материал только онлайн 	<ul style="list-style-type: none"> • Частные пользователи начинается с \$ 5,99 в месяц со сроком хранения в интернете до 6 месяцев. • Организации аккаунты начинаются от \$19,99 в месяц

<p>10.ТПУ антиплагиат</p>	<p>Онлайн</p>	<p>Технология на основе искусственного интеллекта и шаг шингла 4</p>	<ul style="list-style-type: none"> • подробный отчет о плагиате • достаточное количество принимаемых файлов • Размер файла до 20 Мб. • Дополнительные параметры типов файлов. • Подсчет цитируемости и само цитирования 	<ul style="list-style-type: none"> • Количество проверок: не более 3 файлов в течение 24 часов и не более 20 файлов в месяц. • Размер загружаемого файла не должен превышать 100 Мб • Примерное время проверки документа: 14 мин. (зависит от загруженности сервиса) 	<ul style="list-style-type: none"> • бесплатно
-------------------------------	---------------	--	--	---	---

2. РЕАЛИЗАЦИЯ СИСТЕМЫ ПРОВЕРКИ ОРИГИНАЛЬНОСТИ РАБОТ

2.1 Пользовательские требования к разрабатываемой системе

В процессе работы были сформулированы требования пользователя к разрабатываемой системе:

1. Возможность выбора файлов для проверки их сходства при помощи приложения.
2. Возможность загрузки файлов для проверки их сходства при помощи приложения.
3. Возможность выбора базового файла из списка для последующего сравнения других файлов с базовым.
4. Возможность отфильтровать результаты сходства из списка файлов на основе минимального процентного сходства, для последующего анализа результатов.
5. Возможность одновременного вычисления сходства всех выбранных файлов относительно каждого файла в списке в качестве базового файла, чтобы выполнить анализ за один раз.
6. Возможность выбора размера шинглов для алгоритма, для изменения точности алгоритма.
7. Возможность изменения минимального процента отображения для отображения результата сходства, чтобы результаты соответствовали моему желаемому диапазону.
8. Возможность кластеризации результатов сравнения сходства на основе выбранного процента, для визуализации результатов.
9. Возможность удаления нежелательных результатов подобия из отображаемого списка, для визуализации результатов.

2.2 Варианты использования системы

Основной сценарий

1. Пользователь инициирует добавление файлов.

2. Система открывает окно для выбора файлов с ПК.
3. Пользователь выбирает файлы, которые он собирается сравнивать.
4. Система выделяет выбранные файлы.
5. Пользователь подтверждает выбор для загрузки.
6. Система закрывает диалоговое окно.
7. Система загружает и отображает список выбранных файлов.

Альтернативный сценарий

Предусловие: на шаге 5 основного сценария пользователь не подтверждает отказ.

Переход к шагу 3 основного сценария.

Диаграмма вариантов использования системы приведена на рисунке 5.



Рисунок 5 — Диаграмма вариантов использования в нотации UML

2.3 Используемые технологии

Для разработки приложения использовался Flutter, поскольку это бесплатный набор для разработки мобильного пользовательского интерфейса с открытым исходным кодом, созданный Google [1]. В качестве языка программирования был выбран Dart. Дарт имеет следующие преимущества:

- обширная коллекция библиотек и фреймворков;
- подходит для больших проектов благодаря встроенной поддержке сильного шрифта;
- простой и лаконичный синтаксис.

Соответственно, для данного языка программирования была выбрана наиболее подходящая среда разработки VS Code. Он простой инструмент, все конфигурируется через плагины с маркетплейса. В нашем случае настольное приложение более удобно, так как пользователь должен работать в автономном режиме и с уже загруженным файлом на своем локальном компьютере.

2.4 Алгоритм работы программы

По требованию клиента наше приложение должно иметь возможность загружать файлы (.txt и .docx) из папки на локальном диске. Это было сделано с помощью пакета File Picker 4.5.1, который позволяет использовать собственный файловый менеджер для выбора одного или нескольких файлов с поддержкой фильтрации расширений. Затем хеш-карта сопоставляет каждый ключ ровно с одним текстовым документом, чтобы обеспечить итерацию по файлам. Создается список всех загруженных файлов и с помощью созданных ключей они отображаются при первом запуске приложения. К этим файлам применяется алгоритм шингла, и полученные проценты сходства отображаются в окне вкладок результатов. Результаты показывают, насколько другие файлы похожи на выбранный базовый файл из списка. Полная схема приложения показана ниже на рисунке 6 со встроенным в него алгоритмом шинглов (рисунок 7). Более надежным способом использования приложения является одновременное пакетное сравнение всех файлов. Вместо того, чтобы каждый раз вручную выбирать другой базовый файл, процесс автоматизирован и выполняет сравнение за один цикл. Конечным продуктом является исчерпывающий список сходства документов относительно каждого возможного базового файла в папке. На приведенной ниже схеме видно, что алгоритм шинглы будет повторяться n раз (n — общее количество выбранных документов для сравнения) если выбрано пакетное сравнение. Результаты на этом этапе представлены в виде имен файлов

и их сходства с базовым файлом. Пользователь может завершить программу здесь и интерпретировать результаты в этой форме. Однако для большего удобства пользователь может кластеризовать результаты на основе выбранного минимального процента сходства. Процесс кластеризации имеет две подгруппы. К первой группе относятся студенты, у которых содержание отчетов соответствует друг другу выше порога допустимого сходства. Эти результаты представлены парами, поэтому пользователь (преподаватель) может легко определить виновных, скопировавших друг друга. Другая кластеризация — это когда группа отчетов соответствует конкретному базовому отчету/файлу выше допустимого порогового процента, но базовый отчет не соответствует им в такой степени. Это отражает одностороннее копирование этого документа другими учащимися. В таком кластере может быть указано более двух документов.

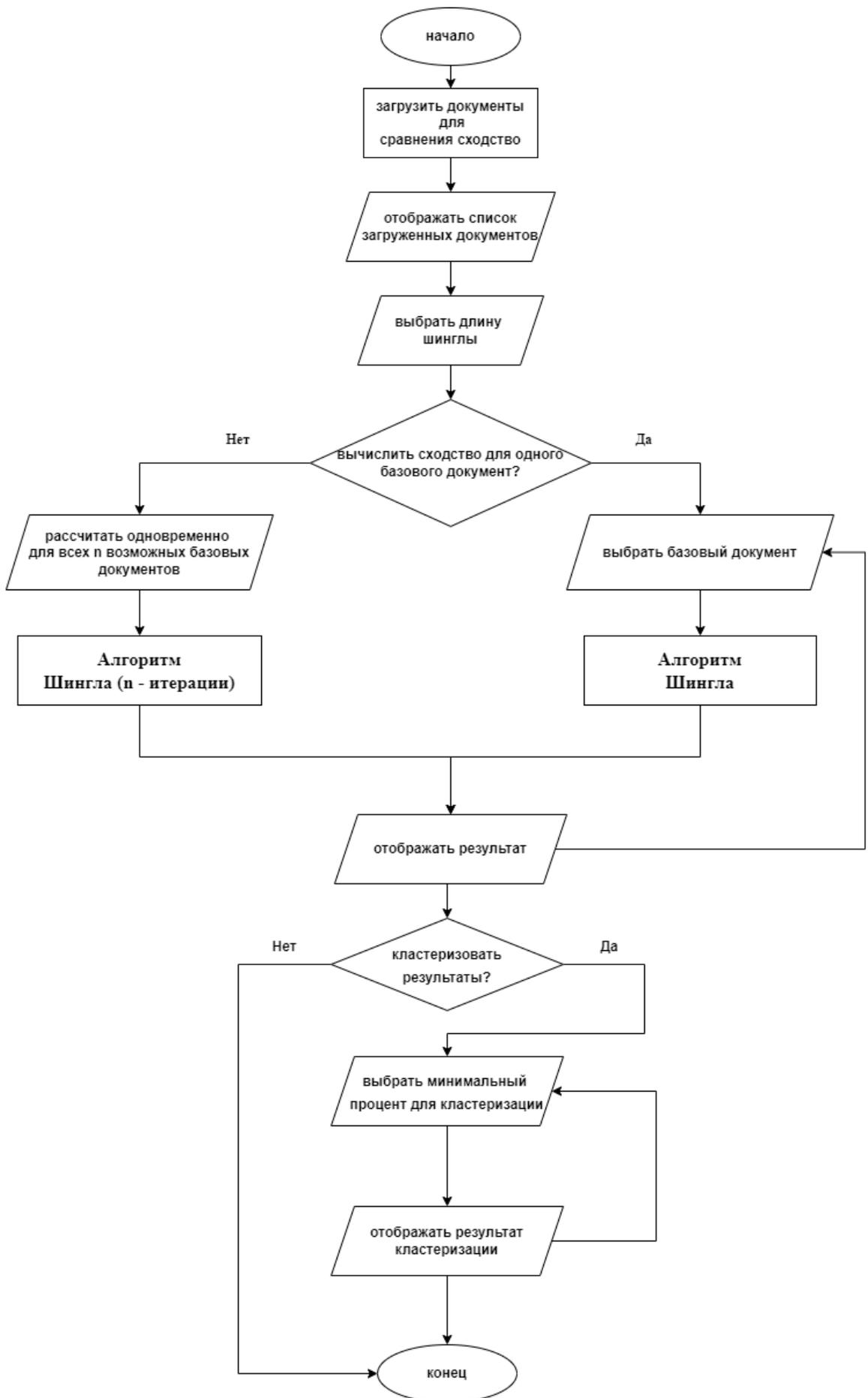


Рисунок 6 — Схема работы приложений

2.5 Алгоритм шинглов

Основная идея алгоритма шинглы состоит в том, чтобы преобразовать проблему подобия файлов в проблему подобия множества. Установленные методы измерения сходства в основном включают сходство (3) и включение (4), которые определяются следующим образом:

$$R(f_1, f_2) = \frac{|shingle(f_1) \cap shingle(f_2)|}{|shingle(f_1) \cup shingle(f_2)|} \quad (3)$$

где

$R(f_1, f_2)$ сходство между файлом F1 и файлом F2;

$$C(f_1, f_2) = \frac{|shingle(f_1) \cap shingle(f_2)|}{|shingle(f_1)|} \quad (4)$$

где

$C(f_1, f_2)$ сходство (содержание) файла F2 в файле F1.

В нашем сценарии более логичная форма сравнения сходства файлов учащегося основана на формуле 4. Мы используем работу одного учащегося в качестве базового файла, а остальные отчеты сравниваем с этим отдельным отчетом. Алгоритм Шинглов имеет большую пространственную и вычислительную сложность, а также высокую точность подобия. Он подходит для приложений с небольшим объемом данных и высокими требованиями к точности. Поэтому с течением времени выбирается компромисс для точности. Для определения схожести двух документов с помощью алгоритм шингла необходимо:

1. канонизация текста :
 - очистка от таких спец символов;
 - очистка от номеров;
 - очистка от ненужных частей речи в предложении;
2. разбиение на шинглы;
 - делим текст по пробелам и создаем массивы;
3. сравнение, определение результата.

- используйте формулу (4) и умножьте результаты на 100%, чтобы получить процентное сходство

Схема алгоритма шинглов показана ниже на рисунке 7.

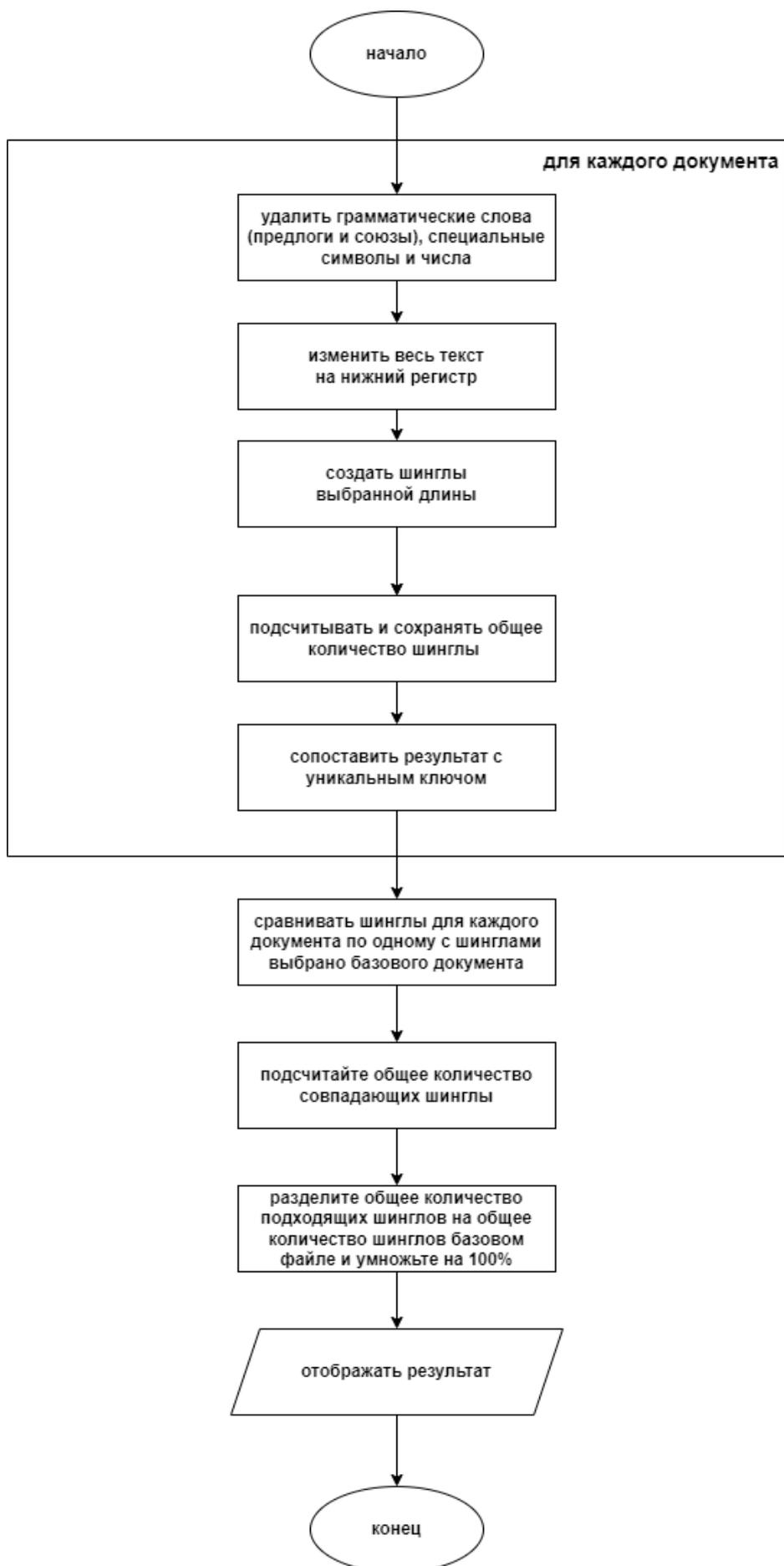


Рисунок 7 — Схема работы алгоритма шинглы

3. РЕЗУЛЬТАТЫ РАБОТЫ СИСТЕМЫ

На рисунках с 8 до 14 ниже показан поэтапный функционал программы:

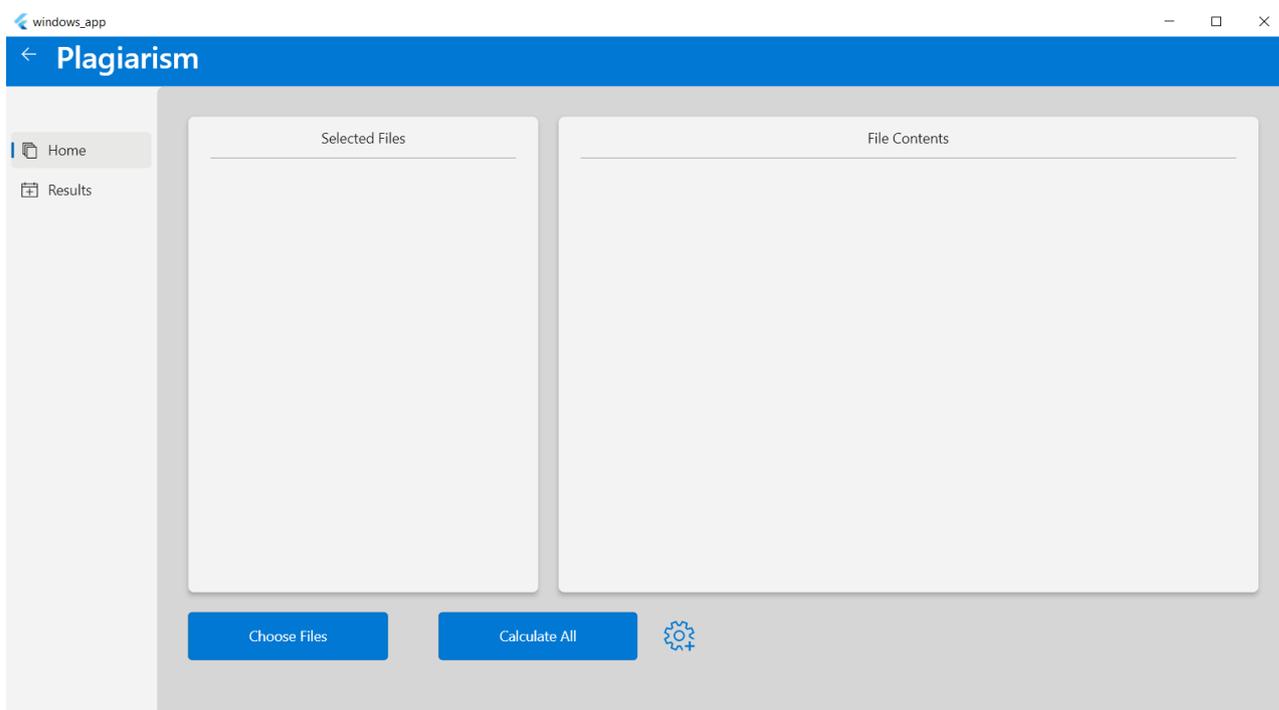


Рисунок 8 — GUI при загрузке программы

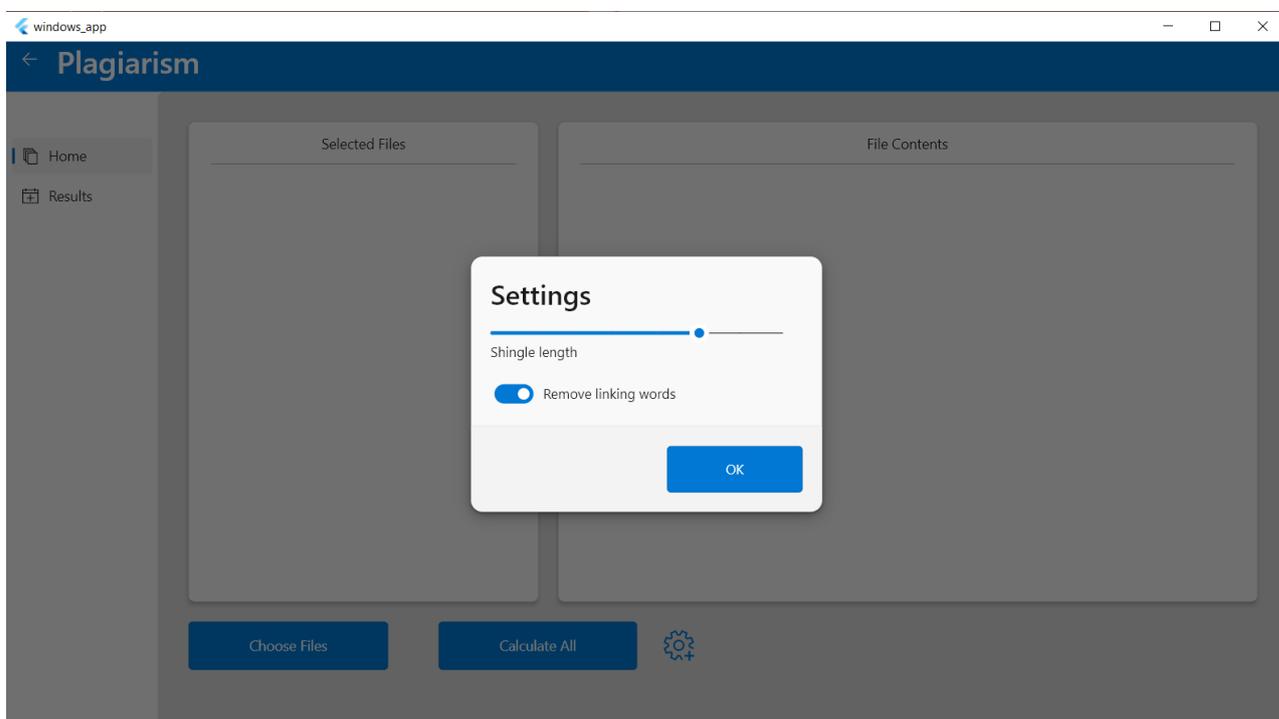


Рисунок 9 — Выбор длины шинглы

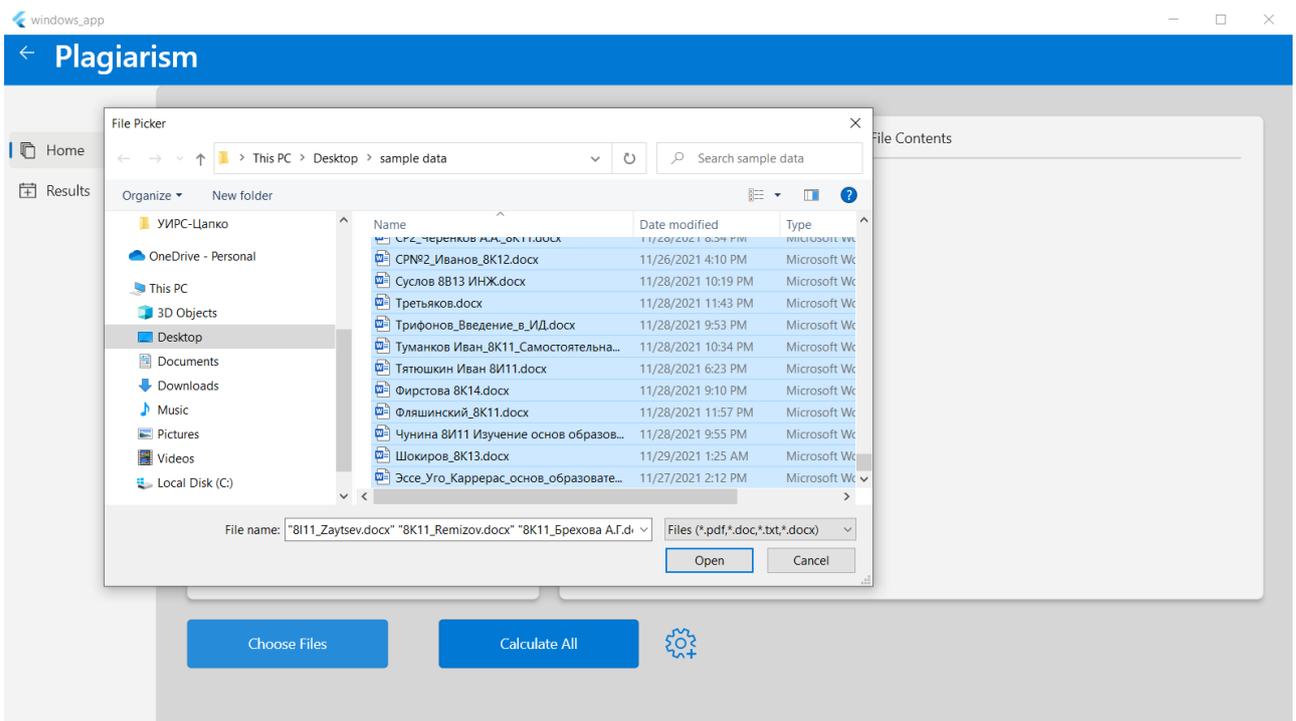


Рисунок 10 — Выбор файлов

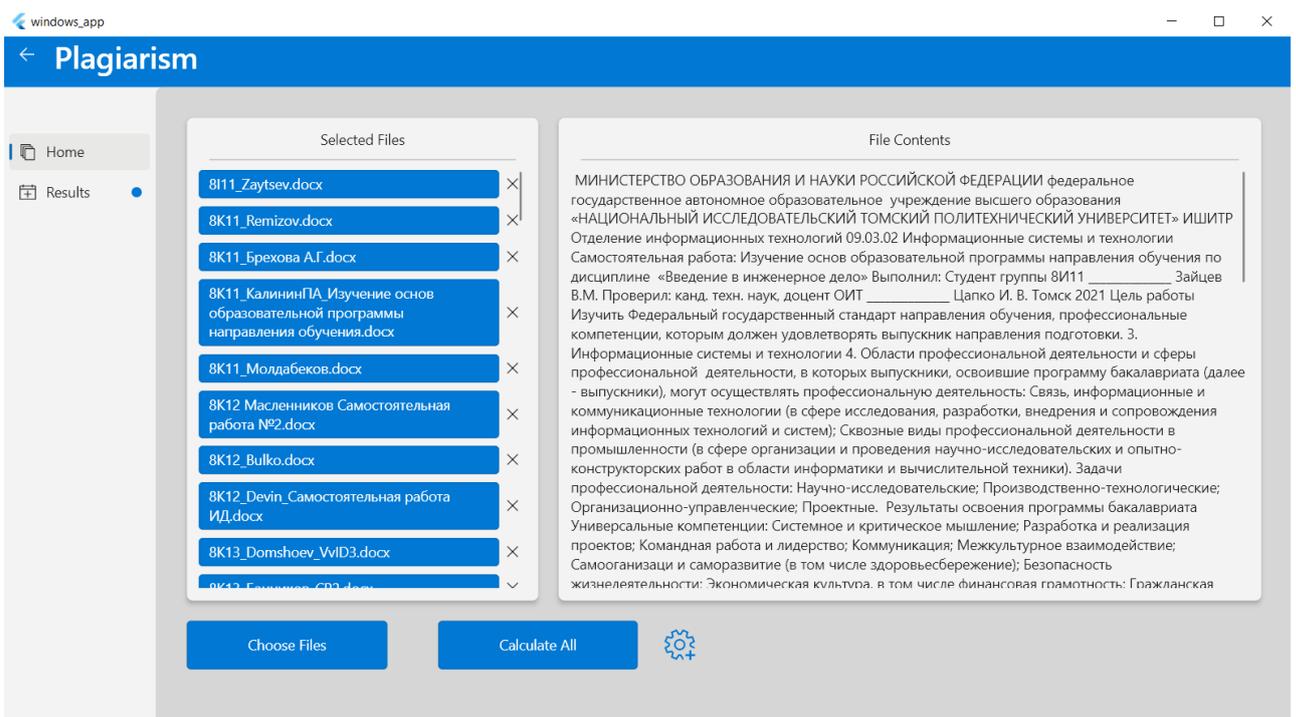


Рисунок 11 — Выбор базового файла (8I11_Zaytsev.docx)

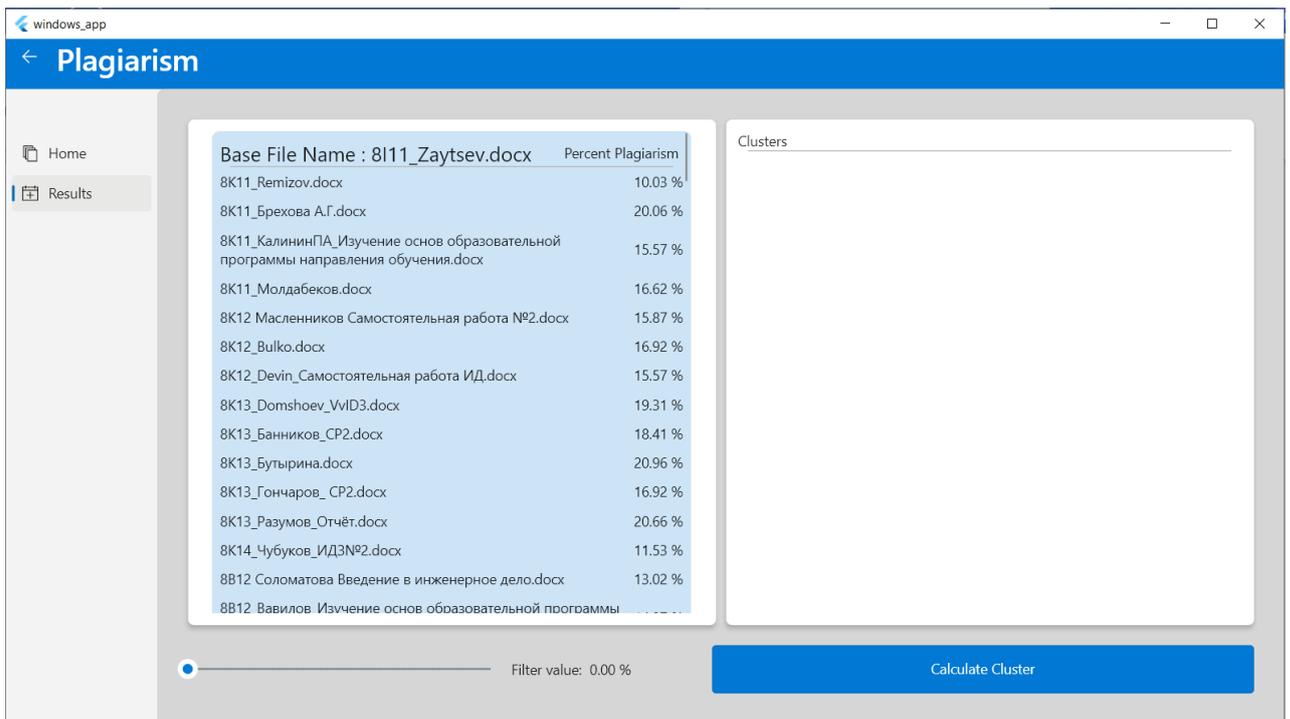


Рисунок 12 — Результаты сходства на базовом файле (8I11_Zaytsev.docx)

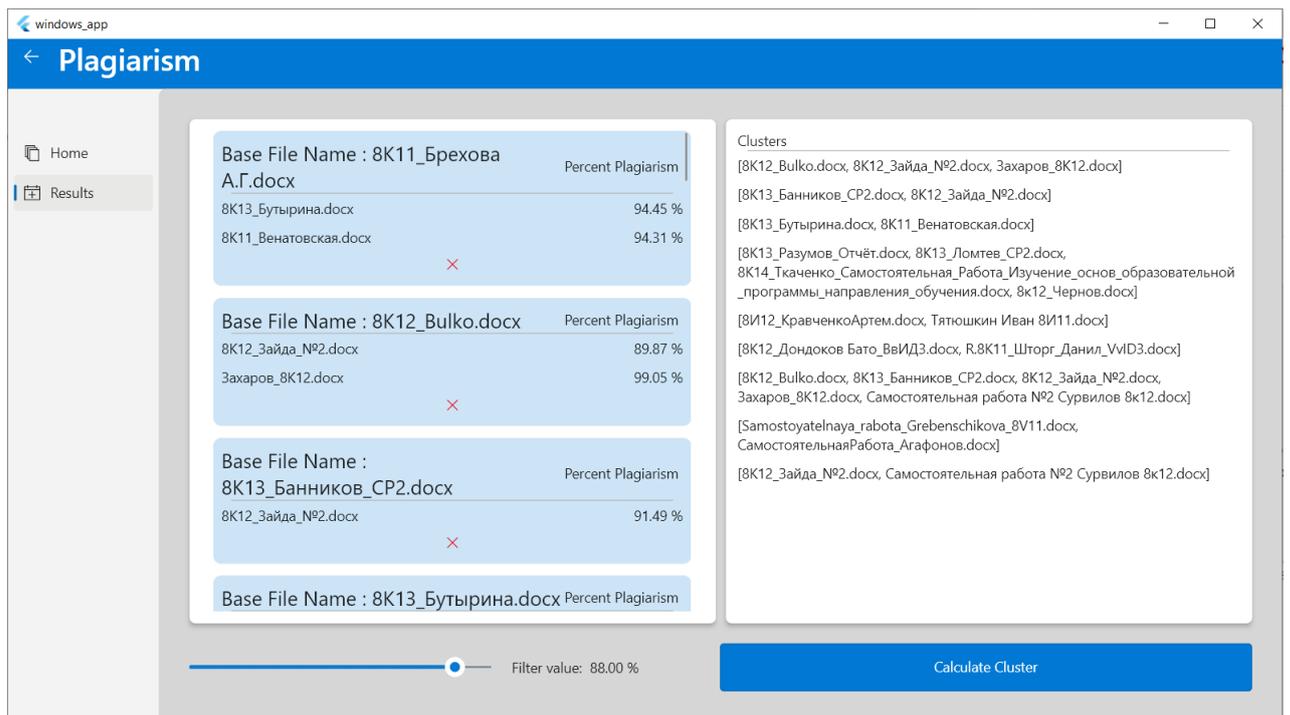


Рисунок 13 — Результаты кластеризации на общем файле

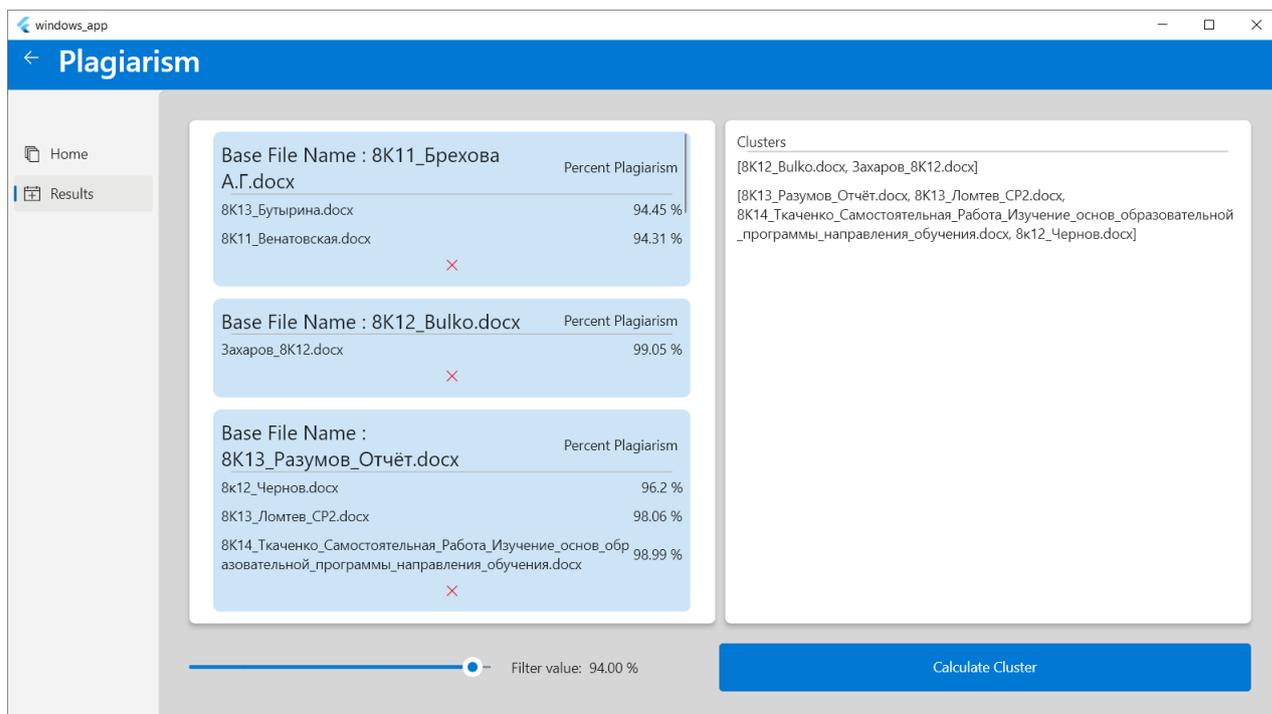


Рисунок 14 — Результаты кластеризации на общем файле

Для сравнения точности реализованного алгоритма шинглов результаты разработанного приложения для борьбы с плагиатом, основанного на файле (8I11_Zaytsev.docx), сравнивались с результатами из Plagiarism CheckerX. Результаты были удовлетворительно точными как показано в таблице 2 Приложение А и не превышали 5% результата, обнаруженного средством проверки на плагиат X. Разница возникает из-за различий в используемых алгоритмах. Кластеризация показала правильные результаты, как и ожидалось от тестовой папки. Более внимательное изучение содержимого этих помеченных файлов показало, что студенты действительно копировали друг друга. Программе требуется 10–11 минут, чтобы одновременно обработать все 183 файла общим размером 9 МБ и отобразить результаты. Это существенный результат, так как текущий Антиплагиат в университете обрабатывает 1 файл за 4 минуты.

4. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Выпускная квалификационная работа посвящена разработке автономной системы проверки оригинальности студенческих работ. Система позволит осуществлять процесс проверки оригинальности студенческих работ (отчетов), представленных преподавателям внутри университетских групп. Большинство существующих систем антиплагиата ориентированы на онлайн-проверку на плагиат опубликованных материалов в глобальных академических базах данных. В этих системах отсутствуют функции проверки оригинальности работы неопубликованных материалов в аудитории, когда студенты представляют их преподавателям для оценки. Это с годами создало культуру повторного использования старых отчетов студентов высших курсов для выполнения заданий. Негативные последствия этой привычки проявляются, когда студенты с трудом справляются с написанием курсовых и выпускных квалификационных работ, требующих проверки на плагиат, поскольку они никогда не овладевали навыками исследования и написания оригинальной работы.

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

4.1.1 Потенциальные потребители результатов исследования

Компания Technavio объявила в своем последнем отчете об исследовании рынка под названием «Рынок программного обеспечения для защиты от плагиата для образовательного сектора по конечным пользователям и географическому положению — прогноз и анализ на 2021–2025 годы», что рынок программного обеспечения для защиты от плагиата вырастет на 1,73 миллиарда долларов США при среднегодовом темпе роста. Ключевой сегмент рынка антиплагиата по данным исследования Technavio показан на рисунке 4.1. Совершенно очевидно, что Россия и Европа имеют самые быстрорастущие рынки и, следовательно, являются наиболее прибыльным целевым рынком с точки зрения региона.

Which are the Key Regions for Anti-plagiarism Software Market for Education Sector?



Рисунок 4.1 – Ключевые регионы рынка антиплагиатного программного обеспечения в образовательной сфере

Сегментировать рынок услуг по разработке офлайн системы проверки оригинальности студенческих работ можно по следующим критериям: конечный потребитель продукции (вид ВУЗа), регион (рисунок 4.2).

		Вид ВУЗа			
		Федеральный университет	Университет	Академия	Институт
Регион	Россия			С	
	Европа		СР		С
	Америка	РСХ	РСХ	СР	РСХ

СР	Check Plagiarism	РСХ	Plagiarism CheckerX
С	Copyleaks		

Рисунок 4.2 – Карта сегментирования рынка услуг по разработке офлайн антиплагиат

Потенциальными потребителями разработки являются сотрудники (преподаватели) Национального исследовательского Томского политехнического университета, вовлеченные в процесс обучения студентов, а

также других высших учебных заведений, желающих повысить исследовательскую культуру и академическую честность в образовании.

4.2 Анализ конкурентных технических решений

Анализ конкурентоспособных технических решений с позиций ресурсоэффективности и ресурсосбережения позволяет оценить сравнительную эффективность научных разработок и определить направления ее дальнейшего совершенствования. Этот анализ был проведен с использованием оценочной карты, сравнивающей конкурирующие системы для проверки сходства документов в автономном режиме, как представлено в таблице 4.1.

Таблица 4.1 – Оценочная карта для сравнения конкурентных офлайн антиплагиат.

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б _Ф	Б ₁	Б ₂	К _Ф	К ₁	К ₂
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Функциональность (в соответствии с потребностями клиентов)	0,1	5	3	2	0,5	0,3	0,2
2. Точность алгоритма	0,15	4	5	3	0,6	0,75	0,45
3. Потребность в ресурсах памяти	0,1	3	4	3	0,3	0,4	0,3
4. Надежность	0,1	4	4	3	0,4	0,4	0,3
5. Безопасность	0,1	5	4	4	0,5	0,4	0,4
6. Скорость работы алгоритмов	0,15	4	5	3	0,6	0,75	0,45
Экономические критерии оценки эффективности							
7. Цена	0,1	5	3	3	0,5	0,3	0,3
8. Конкурентоспособность	0,1	4	4	3	0,4	0,4	0,3
9. Послепродажное обслуживание	0,05	5	3	3	0,25	0,15	0,15
10. Предполагаемый срок эксплуатации	0,05	5	3	4	0,25	0,15	0,2
Итого	1	44	38	31	4,3	4	3,05

Где Б_Ф – Plagiarism (моё приложение);

Б₁ – Plagiarism CheckerX ;

Б₂ – Антиплагиат (текущая система в ТПУ) .

Анализ конкурентных технических решений определяется по формуле:

$$K = \sum B_i \times B_i \quad (4.1)$$

где K – конкурентоспособность вида;

B_i – вес критерия (в долях единицы);

B_i – балл i -го показателя.

Из представленного данных оценочной карты можно сделать вывод, что мое приложение (Plagiarism) предлагает желаемую функциональность по самой низкой цене без ущерба для безопасности и точности по сравнению с его конкурентами Plagiarism CheckerX и Anti-Plagiarism.

4.3 SWOT– анализ

Для анализа сильных и слабых сторон проекта, а также выявления возможностей и угроз со стороны внешней среды был выбран метод SWOT-анализ, позволяющий оценить факторы и явления, способствующие или препятствующие продвижению проекта на рынок. На первом этапе SWOT анализа в таблице 4.2 были описаны сильные и слабые стороны проекта, выявлены возможности и угрозы реализации НИ.

Таблица 4.2 – Матрица SWOT анализа

Сильные стороны	Возможности во внешней среде
<p>С1. Приложение сравнивает неопубликованные документы в автономном режиме, используя функцию массового сравнения, которой нет у большинства антиплагиатов;</p> <p>С2. Используемый алгоритм шинглов очень прост для понимания и достаточно точен;</p> <p>С3. Приложение очень дешевое по сравнению с конкурентами;</p> <p>С4. Наличие опытного руководителя;</p> <p>С5. Приложение очень простое в настройке и не требует подключения к интернету;</p> <p>С6. Для использования не требуется регистрация</p> <p>С7. Легко распространяется на разных платформах (Windows OS, macOS)</p>	<p>В1. Легкая адаптация алгоритма под иностранные языки;</p> <p>В2. Большой потенциал для применения метода в России и других странах;</p> <p>В3. Публикации о проекте в тематических журналах.</p> <p>В4. Может быть дополнительно оптимизирован в будущем для повышения скорости и точности.</p> <p>В5. Легко конвертируется в веб-приложение, если возникает необходимость</p>

Слабые стороны	Угрозы внешней среды
Сл1. Не полностью оптимизирован; Сл2. Неспособность алгоритма пометить семантический плагиат; Сл3. Значительное время начальной загрузки.	У1. Крупные компании могут добавить функцию сравнения в автономном режиме, когда осознают ее потребность; У2. Отсутствие финансирования У3. Недостаточный спрос на офлайн-антиплагиат неопубликованного материала.

Соответствие сильных и слабых сторон исследовательского проекта внешней среде помогает определить степень необходимости стратегических изменений. Каждый фактор отмечен либо знаком «+» (что означает сильное соответствие сильных сторон возможностям), либо знаком «-» (что означает слабое соответствие); "0" – не определен. Интерактивные матрицы проекта представлена в таблицах 4.3 и 4.4:

Таблица 4.3 – Интерактивная матрица сильных и слабых сторон и возможностей

		Сильные стороны							Слабые стороны		
Возможности		C1	C2	C3	C4	C5	C6	C7	Сл1	Сл2	Сл3
Проекта	B1	0	+	+	-	+	+	+	-	-	-
	B2	+	+	+	0	+	+	+	-	-	-
	B3	0	+	-	+	-	-	-	-	-	-
	B4	+	+	-	-	-	0	+	+	-	+
	B5	-	0	+	0	-	-	+	-	-	+

Таблица 4.4 – Интерактивная матрица сильных сторон и слабых сторон и угроз

		Сильные стороны							Слабые стороны		
Угрозы		C1	C2	C3	C4	C5	C6	C7	Сл1	Сл2	Сл3
проекта	У1	+	-	-	-	-	-	-	-	+	-
	У2	-	-	-	-	-	-	-	0	+	+
	У3	+	-	+	-	0	0	-	-	0	-

Анализ интерактивных таблиц представляется в форме записи сильно коррелирующих сильных сторон и возможностей или слабых сторон и возможностей:

- В1С2С3С5С6С7; В2С1С2С3С5С6С7; В3С2С4; В4С1С2С7; В5С3С7;
- В4Сл1Сл3; В5Сл3.

Анализ интерактивных таблиц представляется в форме записи сильно коррелирующих сильных сторон и угрозой или слабых сторон и угрозой:

- У1С1; У3С1С3;
- У1Сл2; У2Сл2Сл3.

Самой большой угрозой для проекта является возможное добавление функции проверки подобия документов в автономном режиме к уже установленным ранее существующим системам антиплагиата, как только они осознают потребность в этой функции. Что касается слабостей, то неспособность алгоритма пометить семантический плагиат делает его уязвимым для обмана путем перефразирования плагиата. В рамках третьего этапа составляется итоговая матрица SWOT-анализа, представленная в таблице 4.5.

Таблица 4.5 – Итоговая матрица SWOT-анализа

	<p>Сильные стороны научно-исследовательского проекта:</p> <p>С1. Приложение сравнивает неопубликованные документы в автономном режиме, используя функцию массового сравнения, которой нет у большинства антиплагиатов;</p> <p>С2. Используемый алгоритм шинглов очень прост для понимания и достаточно точен;</p>	<p>Слабые стороны научно-исследовательского проекта:</p> <p>Сл1. Не полностью оптимизирован;</p> <p>Сл2. Неспособность алгоритма пометить семантический плагиат;</p> <p>Сл3. Значительное время начальной загрузки.</p>
--	---	---

	<p>С3. Приложение очень дешевое по сравнению с конкурентами;</p> <p>С4. Наличие опытного руководителя;</p> <p>С5. Приложение очень простое в настройке и не требует подключения к интернету;</p> <p>С6. Для использования не требуется регистрация</p> <p>С7. Легко распространяется на разных платформах (Windows OS, macOS)</p>	
<p>Возможности:</p> <p>В1. Легкая адаптация алгоритма под иностранные языки;</p> <p>В2. Большой потенциал для применения метода в России и других странах;</p> <p>В3. Публикации о проекте в тематических журналах.</p> <p>В4. Может быть дополнительно оптимизирован в будущем для повышения скорости и точности.</p> <p>В5. Легко конвертируется в веб-приложение, если возникает необходимость</p>	<p>Развитию и совершенствованию приложения способствует большой потенциал применения алгоритма шинглов для обнаружения сходства файлов в университетах России и других стран.</p>	<p>Алгоритм нуждается в дальнейшей оптимизации, а добавление семантического обнаружения плагиата наверняка повысит его конкурентоспособность на рынке.</p>

<p>Угрозы:</p> <p>У1. Крупные компании могут добавить функцию сравнения в автономном режиме, когда осознают ее потребность;</p> <p>У2. Отсутствие финансирования</p> <p>У3. Недостаточный спрос на офлайн-антиплагиат неопубликованного материала.</p>	<p>Угроза со стороны авторитетных компаний, добавляющих этот функционал в собственные программы, позволяет сделать проект глобальным.</p>	<p>Самой большой угрозой для проекта является возможное добавление функции проверки подобия документов в автономном режиме к уже установленным ранее существующим системам антиплагиата, как только они осознают потребность в этой функции.</p>
--	---	--

Исходя из результатов SWOT-анализа, можно сделать вывод, что, несмотря на угрозы и слабые стороны проекта, проект можно считать перспективным и успешным.

4.4 Планирование работ по научно-техническому исследованию

4.4.1 Структура работ в рамках научного исследования

Прежде чем приступить к выполнению задания в рамках научного исследования, необходимо спланировать комплекс предполагаемых работ. Это требуется для эффективного распределения задач во времени, необходимого для затрат на разработку решения и оформления выпускной квалификационной работы. В качестве исполнителей работ выступают студент и научный руководитель. Студент отвечает за поиск литературы, анализ предметной области, проектирование, разработку решения, тестирование. Научный руководитель отвечает за корректировку целей и задач работы, сформированную студентом, проверку выполнения работы и выявление недочетов. Перечень работ и распределение исполнителей представлены в таблице 4.6.

Таблица 4.6 – Перечень этапов, работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Должность исполнителя
	1	Выбор научного руководителя	Бакалавр

Выбор направления исследований	2	Составление и утверждение темы бакалаврской работы	Бакалавр Руководитель
Разработка технического задания	3	Обозначение и утверждение плана работ	Бакалавр
	4	Составление календарного плана-графика выполнения бакалаврской работы	Бакалавр
	5	Подбор и изучение материалов по теме	Бакалавр
	6	Анализ предметной области	Руководитель Бакалавр
Теоретические и экспериментальные исследования	7	Разработка алгоритма шинглов для использования в приложении	Руководитель Бакалавр
	8	Разработка приложения	Бакалавр
Обобщение и оценка результатов	9	Согласование результатов работы с научным руководителем	Руководитель Бакалавр
Оформление отчета по НИР	10	Составление пояснительной записки	Бакалавр

4.4.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования. Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, который зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости $t_{ожi}$ используется следующая формула:

$$t_{ожi} = \frac{3t_{\min i} + 2t_{\max i}}{5}, \quad (4.2)$$

где $t_{ожi}$ – ожидаемая трудоемкость i -ой работы (чел.-дни);

$t_{\min i}$ – минимально возможная трудоемкость выполнения заданной i -ой работы (оптимистическая оценка), (чел.-дни);

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

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях T_p , учитывающая параллельность выполнения работ по нескольким исполнителями.

$$T_{pi} = \frac{t_{ожi}}{Ч_i}, \quad (4.3)$$

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

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

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

4.4.3 Разработка графика проведения научного исследования

Наиболее удобным и наглядным представлением проведения научных работ является построение ленточного графика в форме диаграммы Ганта. Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ. Для удобства построение графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{\text{кал}}, \quad (4.4)$$

где T_{ki} – продолжительность выполнения i -й работы в календарных днях;

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

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

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - (T_{\text{вых}} + T_{\text{пр}})}, \quad (4.5)$$

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

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

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

Расчет коэффициента календарности:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - (T_{\text{вых}} + T_{\text{пр}})} = \frac{365}{365 - 118} = 1,48$$

Таблица 4.7 – Временные показатели проведения научного исследования

Название работы	Трудоёмкость работ						Исполнители	Длительность работ в рабочих днях T_{pi}		Длительность работ в календарных днях T_{ki}	
	T_{min} , чел–дни		T_{max} , чел–дни		$T_{ож}$, чел– дни			Исп.1	Исп.2	Исп.1	Исп.2
	Исп.1	Исп.2	Исп.1	Исп.2	Исп.1	Исп.2					
Выбор научного руководителя	1	-	2	-	1,4	-	Бакалавр	1,4	-	3	-
Составление и утверждение темы бакалаврской работы	2	1	4	2	2,8	1,4	Бакалавр Руководитель	1,4	0,7	3	2
Обозначение и утверждение плана работ	2	-	4	-	2,8	-	Бакалавр	2,8	-	5	-
Составление календарного плана-графика выполнения бакалаврской работы	2	-	4	-	2,8	-	Бакалавр	2,8	-	5	-
Подбор и изучение материалов по теме	15	-	20	-	17	-	Бакалавр	17	-	26	-

Анализ предметной области	5	5	8	8	6,2	6,2	Бакалавр Руководитель	3,1	3,1	5	5
Разработка алгоритма шинглов для использования в приложении	20	2	24	4	21,6	2,8	Бакалавр Руководитель	10,8	1,4	16	3
Разработка приложения	24	-	30	-	26,4	-	Бакалавр	26,4	-	40	-
Согласование результатов работы с научным руководителем	3	2	5	4	3,8	2,8	Бакалавр Руководитель	1,9	1,4	3	3
Составление пояснительной записки	10	-	16	-	12,4	-	Бакалавр	12,4	-	19	-
Итого	80	10	117	18	94,8	13,2		80	6,6	125	13

На основе рассчитанных временных показателей проведения научного исследования была построена диаграмма Ганта на таблице 4.8 и визуализирована занятость ресурсов на таблице 4.9.

Таблица 4.8 Календарный план-график проведения научного исследования

№ работ	Вид работ	Исполнители	Т _{ки} , кал.дн.	Продолжительность выполнения работ													
				февраль		март			апрель			май					
				2	3	1	2	3	1	2	3	1	2	3			
1	Выбор научного руководителя	Бакалавр	3	■													
2	Составление и утверждение темы бакалаврской работы	Бакалавр Руководитель	3	■ ■													
3	Обозначение и утверждение плана работ	Бакалавр	5		■												
4	Составление календарного плана-графика выполнения бакалаврской работы	Бакалавр	5			■											
5	Подбор и изучение материалов по теме	Бакалавр	26				■	■	■	■							
6	Анализ предметной область	Бакалавр Руководитель	5								■ ■						

4.5 Бюджет научно-технического исследования (НТИ)

1. Материальные затраты.
2. Затраты на спец.оборудование.
3. Основная и дополнительная ЗП.
4. Социальные отчисления.
5. Прямые затраты.
6. Накладные расходы.

4.5.1 Расчет материальных затрат НТИ

При планировании бюджета научно-техническое исследование должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. Расчет материальных затрат осуществляется по формуле:

$$Z_M = (1 + k_T) \cdot \sum_{i=1}^m C_i \cdot N_{расхi}, \quad (4.6)$$

где k_T – коэффициент, учитывающий транспортно-заготовительные расходы;

m – количество видов материальных ресурсов, потребляемых при выполнении научного исследования;

C_i – цена приобретения единицы i -го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, и т. д.);

$N_{расхi}$ – количество материальных ресурсов i -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, и т. д.).

Таблица 4.9 – Материальные затраты

Наименование	Единица измерения	Количество			Цена за ед., руб.	Затраты на материалы, (Z_M), руб.		
		Исп.1	Исп.2	Исп.3		Исп.1	Исп.2	Исп.3
Тетрадь для записей	Шт.	1	1	1	50	50	50	50
Ручка	Шт.	1	1	1	20	20	20	20
Электроэнергия	кВт*ч	250	300	280	5	1250	1500	1400
Итого, руб.						1320	1570	1470

4.5.2 Расчет затрат на специальное оборудование для научных работ

Все расчеты по приобретению спецоборудования и оборудования, имеющегося в организации, но используемого для каждого исполнения конкретной темы, сводятся в таблице 4.10.

Таблица 4.10 – Расчет бюджета затрат на приобретение спецоборудования для научных работ

Наименование	Количество единиц оборудования			Цена единицы оборудования, тыс. руб.			Общая стоимость оборудования, тыс. руб.		
	Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3
Персональный компьютер	1	1	1	100			100	100	100
Программное обеспечение (Android Studio лицензия разработчика)	0	1	1	1,5			0	1,5	1,5
Программное обеспечение (Github Basic, Team, Enterprise)	1	1	1	0	2,64	13,86	0	2,64	13,86
Итого:							100	104,14	115,36

4.5.3 Основная заработная плата исполнителя темы

Данная статья затрат включает основную заработную плату, премии и доплаты всех исполнителей проекта. В качестве исполнителей проекта выступают студенты и научный руководитель. Заработная плата рассчитывается по формуле 4.7:

$$Z_{зп} = Z_{осн} + Z_{доп} \quad (4.7)$$

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

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

$Z_{доп}$ – дополнительная заработная плата (12–20 % от $Z_{осн}$).

Основная заработная плата руководителя рассчитывается по следующей формуле 4.8:

$$Z_{\text{осн}} = Z_{\text{дн}} * T_p * (1 + K_{\text{пр}} + K_{\text{д}}) * K_p \quad (4.8)$$

где $Z_{\text{осн}}$ – основная заработная плата одного работника;

$Z_{\text{дн}}$ – среднедневная заработная плата работника, руб.

T_p – продолжительность работ, выполняемых научно-техническим работником, раб. дн.;

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

$K_{\text{д}}$ – коэффициент доплат и надбавок (0,2–0,5);

K_p – районный коэффициент (для Томска – 1,3).

Среднедневная заработная плата рассчитывается по формуле 4.9:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_d} \quad (4.9)$$

где Z_m – месячный должностной оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб. дня $M = 11,2$ месяца, 5–дневная неделя;

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

Должностные оклады исполнителей проекта согласно приказу ТПУ представлены в таблице 4.11.

Таблица 4.11 – Месячные должностные оклады исполнителей

Исполнитель	Районный коэффициент (для Томска)	Размер месячного должностного оклада без учета коэффициента, рубли
Научный руководитель (должность – доцент, степень – кандидат технических наук)	1,3	33664
Студент (ассистент, без степени)	1,3	21760

Баланс рабочего времени для 5 – дневной рабочей недели сформирован в таблицу 4.12.

Таблица 4.12 – Баланс рабочего времени

Показатели рабочего времени	Руководитель	Студент (Бакалавр)
Календарное число дней	365	365
Количество нерабочих дней - выходные дни - праздничные дни	118	118
Потери рабочего времени - отпуск - невыходы по болезни	48 0	48 0
Действительный годовой фонд рабочего времени	199	199

На основе формулы 4.8 и таблиц 4.11–4.12 была рассчитана среднедневная заработная плата:

$$Z_{\text{дн}}(\text{бакалавр}) = \frac{21760 \cdot 11,2}{199} = 1224,68 \text{ рубля} \quad (4.10)$$

$$Z_{\text{дн}}(\text{руководитель}) = \frac{33664 \cdot 11,2}{199} = 1894,66 \text{ рубля} \quad (4.11)$$

Расчет затрат на основную заработную плату приведен в таблице 4.13.

Таблица 4.13 – Расчет основной заработной платы

Исполнители	$Z_{\text{дн}}$, руб.	$k_{\text{пр}}$	$k_{\text{д}}$	$k_{\text{р}}$	$T_{\text{р}}$, раб.дн.	$Z_{\text{осн}}$, руб.
Руководитель	1894,66	0,3	0,4	1,3	13	54 433,59
Бакалавр	1224,68	0,3	0,4	1,3	125	298515.75
Итого						352949.34

Итоговая сумма затрат на основную заработную плату составила 352949,34 руб.

4.5.4 Расчет дополнительной заработной платы

Дополнительная заработная плата учитывает величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска).

Расчет дополнительной заработной платы рассчитывается по формуле 4.12:

$$Z_{\text{доп}} = k_{\text{доп}} \cdot Z_{\text{осн}}, \quad (4.12)$$

где $k_{\text{доп}}$ – коэффициент дополнительной заработной платы, принятый на стадии проектирования за 0,15.

Таблица 4.14 – Расчет дополнительной заработной платы

Исполнители	$Z_{\text{осн}}$, руб.	$K_{\text{доп}}$	$Z_{\text{доп}}$, руб.
Руководитель	54 433,59	0,12	6532,03
Бакалавр	298515,75	0,12	35821,89
Итого			42353,92

Итоговая сумма затрат на дополнительную заработную плату составила 42353,92 руб.

4.5.5 Отчисления во внебюджетные фонды

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 4.13.:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}) \quad (4.13)$$

где $k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

В соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30%. Отчисления во внебюджетные фонды представлены в таблице 4.15.

Таблица 4.15 – Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата, руб.			Дополнительная заработная плата, руб.		
	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3
Руководитель	54 433,59	60 000	65000	6532,03	7200	7800
Студент (Бакалавр)	298515.75	300000	320000	35821,89	36000	38400
Коэффициент отчислений во внебюджетные фонды	0,3					
Итого						
Исполнение 1	118590,98					
Исполнение 2	120960,00					
Исполнение 3	129360,00					

Итоговая сумма отчислений во внебюджетные фонды составила 118590,98 руб.

4.5.6 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов. Их величина определяется по формуле 4.14:

$$Z_{\text{накл}} = (\sum \text{статей}) \cdot k_{\text{нр}} \quad (4.14)$$

где $k_{\text{нр}}$ – коэффициент, учитывающий накладные расходы.

Величину коэффициента накладных расходов можно взять в размере 16% от суммы затрат, подсчитанных выше. Расчет накладных расходов приведен в таблице 4.16.

Таблица 4.16 – Расчет накладных расходов

Статьи затрат	Сумма, руб.		
	Исп. 1	Исп. 2	Исп. 3
Материальные затраты	1320,00	1570,00	1470,00
Затраты на специальное оборудование	100000,00	104140,00	115360,00
Затраты на основную заработную плату	352949,34	360000,00	385000,00
Затраты на дополнительную заработную плату исполнителям проекта	42353,92	43200,00	46200,00
Затраты на отчисления во внебюджетные фонды	118590,98	120960,00	129360,00
Коэффициент накладных расходов	0,16		
Накладные расходы	98434,28	100779,20	108382,40

Итоговая сумма накладных расходов составила 98434,28 руб.

4.5.7 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно–исследовательской работы является основой для формирования бюджета затрат проекта. Определение бюджета затрат на научно–исследовательский проект приведено в таблице 4.17.

Таблица 4.17 – Расчет бюджета затрат НИИ

Наименование статьи	Сумма, руб.			Примечание
	Исп.1	Исп.2	Исп.3	
1. Материальные затраты НИИ	1320,00	1570,00	1470,00	Пункт 4.5.1
2. Затраты на специальное оборудование для научных (экспериментальных) работ	100000,00	104140,00	115360,00	Пункт 4.5.2
3. Затраты по основной заработной плате исполнителей темы	352949,34	360000,00	385000,00	Пункт 4.5.3

4. Затраты по дополнительной заработной плате исполнителей темы	42353,92	43200,00	46200,00	Пункт 4.5.4
5. Отчисления во внебюджетные фонды	118590,98	120960,00	129360,00	Пункт 4.5.5
6. Затраты на научные и производственные командировки	-	-	-	Отсутствуют
7. Контрагентские расходы	-	-	-	Отсутствуют
8. Накладные расходы	98434,28	100779,20	108382,40	Пункт 4.5.6
9. Бюджет затрат НТИ	713648,52	730649,20	785772,40	

Таким образом, общий бюджет НТИ составляет 713648,52 рубля.

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

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования определяется как:

$$I_{\text{фин.р}}^{\text{исп.}i} = \frac{\Phi_{ri}}{\Phi_{\text{max}}} \quad (4.15)$$

где $I_{\text{фин.р}}^{\text{исп.}i}$ – интегральный финансовый показатель разработки;

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

Φ_{max} – максимальная стоимость исполнения научно-

исследовательского проекта.

$$I_{\text{фин.р}}^{\text{исп.1}} = \frac{713648,52}{785772,40} = 0,908;$$

$$I_{\text{фин.р}}^{\text{исп.2}} = \frac{730649,20}{785772,40} = 0,93;$$

$$I_{\text{фин.р}}^{\text{исп3}} = \frac{785772,40}{785772,40} = 1$$

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum_{i=1}^n a_i \times b_i \quad (4.16)$$

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

a_i – весовой коэффициент i -го варианта исполнения разработки;

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

n – число параметров сравнения.

Таблица 4.18 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ Объект исследования	Весовой коэффициент параметра	Исп.1	Исп.2	Исп.3
1. Способствует росту производительности труда пользователя	0,1	5	3	4
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,2	5	4	3
3. Помехоустойчивость	0,15	3	4	3
4. Энергосбережение	0,15	5	3	4
5. Надежность	0,2	5	4	4
6. Материалоемкость	0,2	5	3	3
Итого	1	4,7	3,55	3,45

$$I_{p-\text{исп1}} = 0,1 \cdot 5 + 0,2 \cdot 5 + 0,15 \cdot 3 + 0,15 \cdot 5 + 0,2 \cdot 5 + 0,2 \cdot 5 = 4,7;$$

$$I_{p-исп2} = 0,1 \cdot 3 + 0,2 \cdot 4 + 0,15 \cdot 4 + 0,15 \cdot 3 + 0,2 \cdot 4 + 0,2 \cdot 3 = 3,55;$$

$$I_{p-исп3} = 0,1 \cdot 4 + 0,2 \cdot 3 + 0,15 \cdot 3 + 0,15 \cdot 4 + 0,2 \cdot 4 + 0,2 \cdot 3 = 3,45.$$

Интегральный показатель эффективности вариантов исполнения разработки ($I_{испi}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле 4.17:

$$I_{испi} = \frac{I_{p-испi}}{I_{фин.р}^{испi}} \quad (4.17)$$

$$I_{исп1} = \frac{I_{p-исп1}}{I_{фин.р}^{исп1}} = \frac{4,7}{0,908} = 5,18$$

$$I_{исп2} = \frac{I_{p-исп2}}{I_{фин.р}^{исп2}} = \frac{3,55}{0,93} = 3,82;$$

$$I_{исп3} = \frac{I_{p-исп3}}{I_{фин.р}^{исп3}} = \frac{3,45}{1} = 3,45.$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки показал, что сравнительную эффективность наша проекта наиболее целесообразный вариант из предложенных.

Формула сравнительная эффективность проекта ($\mathcal{E}_{ср}$) показано ниже:

$$\mathcal{E}_{ср} = \frac{I_{исп2}}{I_{исп1}} \quad (4.18)$$

Сравнив значения интегральных показателей эффективности в таблице 4.19, можно сделать вывод, что реализация технологии в первом исполнении является более эффективным вариантом решения задачи, поставленной в данной работе с позиции финансовой и ресурсной эффективности.

Таблица 4.19 – Сравнительная эффективность разработки

№	Показатели	Исп.1	Исп.2	Исп.3
1	Интегральный финансовый показатель разработки	0,908	0,93	1
2	Интегральный показатель ресурсоэффективности разработки	4,7	3,55	3,45
3	Интегральный показатель эффективности	5,18	3,82	3,45
4	Сравнительная эффективность вариантов исполнения	1	0,74	0,67

4.7 Вывод на раздел

В результате выполнения целей раздела можно сделать следующие выводы:

- результатом анализа конкурентных технических решений является выбор одного из вариантов реализации НИР как наиболее подходящего и оптимального по сравнению с другими;

- в ходе планирования для руководителя и студент (бакалавра) был разработан график реализации этапа работ, который позволяет оценивать и планировать рабочее время исполнителей. Определено следующее: общее количество дней для выполнения работ составляет 125 дней;

- для оценки затрат на реализацию проекта разработан проектный бюджет, который составляет 713648,52 руб.

Результат оценки эффективности ИР показывает следующие выводы:

- значение интегрального финансового показателя ИР составляет 0,908, что является показателем того, что ИР является финансово выгодной по сравнению с аналогами;

- значение интегрального показателя ресурсоэффективности ИР составляет 4,7, по сравнению с 3,55 и 3,45;

- значение интегрального показателя эффективности ИР составляет 5,18, по сравнению с 3,82 и 3,45, и является наиболее высоким, что означает, что техническое решение, рассматриваемое в ИР, является наиболее эффективным вариантом исполнения.

5. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

5.1 Введение

Целью выпускной квалификационной работы является разработка системы, предназначена для выявления подобных работ студентов, расположенных на локальном диске, с указанием степени их оригинальности и последующим выделением совпадающих работ в отдельные кластеры. В качестве последнего шага пользователь сможет определить по отображаемым результатам, какие студентов скопировали работы друг друга и в какой степени в процентах. Рабочая зона для этого система представляет собой место оператора ПЭВМ в высших учебных заведениях.

Необходимость проверки уникальности содержания документов является центральной проблемой во многих профессиональных кругах. В академических учреждениях это позволяет поддерживать высокие стандарты исследовательской работы и заданий, представляемых студентами. Улучшение обнаружения плагиата повысит и облегчит образование, которое отражается на качестве учащихся как результат образовательного процесса. Современные антиплагиат системы проверяют на сходство файлы в Базах опубликованных работ в режиме онлайн. Однако в вузах чрезмерное копирование офлайн среди студентов происходит во время выполнения заданий и делает весь образовательный процесс непродуктивным.

Взаимодействия пользователя с системой будет производиться через ПК в офисе, следовательно, в разделе будут рассмотрены правовые, организационные и производственные вопросы обеспечения безопасности человека при работе с системой. Будут выявлены и проанализированы потенциальные вредные и опасные факторы, возникающие в ходе разработки и использования конечного решения, а также будут предложены мероприятия по их предотвращению. Будут рассмотрены вопросы негативного влияния на окружающую среду и возникновения потенциальных чрезвычайных ситуаций на рабочем месте при работе с системой.

Описание рабочей зоны (рабочего места) при эксплуатации:

- область применения – высшие учебные заведения;
- рабочая зона – место оператора ПЭВМ в офисе размеры 4*6 м;
- количество и наименование оборудования рабочей зоны – 1 ПЭВМ;
- рабочие процессы, связанные с объектом исследования,

осуществляющиеся в рабочей зоне – загрузка файлов для проверки схожести документов и контроль параметров процесса кластеризации.

5.2 Правовые и организационные вопросы обеспечения безопасности

Условия труда в рабочем месте регламентируются Трудовым кодексом Российской Федерации от 30.12.2001 N 197-ФЗ (далее ТК РФ), имеющим приоритетное значение перед другими действующими федеральными законами, связанными с трудовыми отношениями [18]. В соответствии со статьей 92 ТК РФ режим рабочего времени работника:

- для работников в возрасте до шестнадцати лет - не более 24 часов в неделю;
- для работников в возрасте от шестнадцати до восемнадцати лет - не более 35 часов в неделю;
- для работников, являющихся инвалидами I или II группы, - не более 35 часов в неделю.

В течение смены работнику должен выделяться перерыв для отдыха и питания продолжительностью не менее 30 минут и не более двух часов. Работодатель обязан предоставлять работнику ежегодный основной оплачиваемый отпуск длительностью в 28 календарных дней. Режим труда и отдыха предусматривает соблюдение определенной длительности непрерывной работы на персональном компьютере (ПК) и перерывов, регламентированных с учетом продолжительности рабочей смены, видов и категории трудовой деятельности.

Вид трудовой деятельности на персональном компьютере в рамках данной работы соответствует группе В – творческая работа в режиме диалога с ПК, категория трудовой деятельности – III (до 6 часов непосредственной работы на ПК). При 8-часовой рабочей смене и работе на ПК, соответствующей описанным выше критериям необходимо через 1,5 – 2,0 часа от начала рабочей смены и через 1,5 – 2,0 часа после обеденного перерыва устраивать регламентированные перерывы продолжительностью 20 минут каждый или продолжительностью 15 минут через каждый час работы. Продолжительность непрерывной работы на ПК без регламентированного перерыва не должна превышать 2 часа. Эффективными являются нерегламентированные перерывы (микропаузы) длительностью 1–3 минуты для выполнения комплекса упражнений. Регламентированные перерывы и микропаузы целесообразно использовать для выполнения комплекса упражнений и гимнастики для глаз, пальцев рук, а также массажа. Комплексы упражнений целесообразно менять через 2–3 недели.

Продолжительность работы (смены) в ночное время сокращается на один час без последующей отработки. К работе в ночное время не допускаются: беременные женщины; работники, не достигшие возраста 18 лет, за исключением лиц, участвующих в создании и (или) исполнении художественных произведений, и других категорий работников в соответствии с настоящим Кодексом и иными федеральными законами. В течение рабочего дня (смены) работнику должен быть предоставлен перерыв для отдыха и питания. Время предоставления перерыва и его конкретная продолжительность устанавливаются правилами внутреннего трудового распорядка или по соглашению между работником и работодателем. Всем работникам предоставляются выходные дни (еженедельный непрерывный отдых). Законодательством РФ запрещена дискриминация по любым признакам и принудительный труд [18].

Если пользователь постоянно загружен работой с ЭВМ, приемлемой является поза сидя. В положении сидя основная нагрузка падает на мышцы, поддерживающие позвоночный столб и голову. В связи с этим при длительном

сидении время от времени необходимо сменять фиксированные рабочие позы. Исходя из общих принципов организации рабочего места, в нормативно-методических документах сформулированы требования к конструкции рабочего места. Условия креслу в рабочем месте регламентируются ГОСТ 21889–76 Система «Человек-машина». Кресло человека-оператора. Кресло должно обеспечивать человеку-оператору соответствующую характеру, условиям труда физиологически рациональную рабочую позу и длительное поддержание основной рабочей позы в процессе трудовой деятельности. По степени подвижности по отношению к средствам отображения информации и органам управления кресла подразделяют на:

- фиксированные;
- подвижные по опорной поверхности;
- подвижные по направляющим в горизонтальной плоскости с обеспечением фиксации в заданном положении;
- вращающиеся вокруг вертикальной оси опорной конструкции с обеспечением фиксации в заданном положении.

Согласно ГОСТ 12.2.032–78 «Рабочее место при выполнении работ сидя» высота рабочей поверхности при организации рабочего места для работ с ПК должна быть равна 655 мм для мужчин и женщин. Кресло пользователя ПК должно создавать условия для поддержания корпуса человека в физиологически рациональном положении и не затруднять рабочих движений; должна регулироваться высота поверхности сиденья и угол наклона спинки. Из-за частого обращения к монитору ПК, его следует располагать под углом $\pm 15^\circ$ от нормальной линии взгляда в вертикальной плоскости и под углом $\pm 15^\circ$ от сагиттальной плоскости в горизонтальной плоскости.

5.3 Производственная безопасность

Таблица 5.1– Вредные и опасные факторы при работе за ПК

Факторы (ГОСТ 12.0.003–2015)	Нормативные документы
---------------------------------	-----------------------

1. Отсутствие или недостатки необходимого искусственного освещения;	ГОСТ 12.4.011–89 Система стандартов безопасности труда (ССБТ). Средства защиты работающих. Общие требования и классификация [19].
2. Монотонность труда, вызывающая монотонию;	
3. Длительность сосредоточенного наблюдения.	
4. Опасные и вредные производственные факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий.	<ul style="list-style-type: none"> • ГОСТ 12.4.011–89 Система стандартов безопасности труда (ССБТ). Средства защиты работающих. Общие требования и классификация [19]. • ГОСТ Р 12.1.019–2017 Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты [20].

5.4 Анализ выявленных вредных и опасных факторов

5.4.1 Отсутствие или недостаток необходимого искусственного освещения

Недостаточная искусственная освещенность рабочей зоны также является одним из важнейших потенциально вредных и опасных факторов. Работа при недостаточном освещении приводит к появлению усталости глаз, головным болям и переутомлению, снижается производительность труда, а при продолжительном воздействии может привести к снижению зрительной работоспособности. Согласно СП 52.13330.2016 «Естественное и искусственное освещение», характеристику зрительной работы за ПК можно отнести к различению объектов при фиксированной и нефиксированной линии зрения средней точности, при которой наименьший или эквивалентный размер объекта различения (буквы и символы на дисплее ПК) составляет более 0,5 мм [21]. Требования к искусственному и естественному освещению представлены в таблице 5.2.

Таблице 5.2 – Требования к искусственному и естественному освещению

помещений

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подразряд зрительной работы	Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %	Искусственное освещение				Естественное освещение	
					освещенность на рабочей поверхности от системы общего освещения, лк	цилиндрическая освещенность, лк	объединенный показатель UGR, не более	коэффициент пульсации освещенности Кп, %, не более	КЕО e_n , %, при	
									верхнем или комбинированном	боковом
Различение объектов при фиксированной и нефиксированной линии зрения: - средней точности	Более 0,5	В	1	Не менее 70	150	-	24	20	2,0	0,5
			2	Менее 70	100	Не регламентируется	24	20	2,0	0,5

В качестве источников искусственного освещения должны быть использованы люминесцентные лампы, лампы накаливания для местного освещения.

5.4.2 Монотонность труда, вызывающая монотонию

На рабочей месте требуется от человека длительного выполнения однообразных действий или непрерывной и устойчивой концентрации внимания в условиях дефицита сенсорных нагрузок. Психические состояния, возникающих при выполнении монотонной работы: психического насыщения (пресыщения), скуки, сонливости, пониженной бдительности. При 8-часовой рабочей смене и работе на ПК, необходимо через 1,5 – 2,0 часа от начала рабочей смены и через 1,5 – 2,0 часа после обеденного перерыва устраивать регламентированные перерывы продолжительностью 20 минут каждый или продолжительностью 15 минут через каждый час работы согласно с ТК РФ.

5.4.3 Длительность сосредоточенного наблюдения

Анализ результатов программы требует наблюдения в течение значительного периода времени, учитывая количество ежедневно сдаваемых студенческих работ. Это в долгосрочной перспективе увеличивает продолжительность сфокусированного наблюдения, что приводит к сенсорной нагрузке. ГОСТ 12.0.003–2015 ССБТ «Опасные и вредные производственные

факторы» рекомендует использовать защитные очки для фильтрации света и снижения нагрузки на глаза. Регулярные перерывы, рекомендованные в ТК РФ, также помогают свести к минимуму вероятность сенсорной нагрузки. Регламентированные перерывы и микропаузы целесообразно использовать для выполнения комплекса упражнений и гимнастики для глаз, пальцев рук, а также массажа. Комплексы упражнений целесообразно менять через 2–3 недели.

5.4.4 Опасные и вредные производственные факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий.

При взаимодействии человека с ПК следует учесть вероятность поражения электрическим током, контакт с которым может привести к электротравме, а в тяжелых случаях – к гибели человека. Поражение электрическим током может произойти вследствие прикосновения к открытым токоведущим частям, находящимся под напряжением, из-за плохой изоляции токоведущих частей компьютера, при работе за ПК влажными руками. Поэтому особенно важно обеспечить пользователя ПК электробезопасностью. Согласно ГОСТ 12.1.038–82 ССБТ. «Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов» значения напряжения прикосновения и токи при работе с ПК должны быть не выше значений, указанных в таблице 5.3 [22].

Таблица 5.3 – Предельно допустимые напряжения прикосновения и токи

Род тока	U, В	I, мА
	не более	
Переменный, 50 Гц	2,0	0,3
Переменный, 400 Гц	3,0	0,4
Постоянный	8,0	1,0

Для обеспечения пользователя ПК электробезопасностью необходимо установить дополнительные оградительные устройства, обеспечивающие недоступность токоведущих частей для прикосновения, обеспечить защитное

заземления или зануления (защитного отключения) электрооборудования. Перед работой с ПК необходимо убедиться в целостности вилки и провода электропитания, в отсутствии видимых повреждений аппаратуры. При работе с ПК запрещается прикасаться к задней панели системного блока и переключать разъемы периферийных устройств работающего устройства.

5.5 Экологическая безопасность

Использование системы конечным пользователем также необходимо рассмотреть с точки зрения экологической безопасности. Во время использования системы загрязнение атмосферы, гидросферы и селитебную зону не происходит, так как отсутствуют выбросы. Однако стоит отметить негативное воздействие использования системы на литосферу за счет утилизации отходов электрооборудования по причине поломок или из-за несоответствия производственным требованиям по причине технологического устаревания. В соответствии с приказом Минприроды России от 30.09.2011 N 792 «Об утверждении Порядка ведения государственного кадастра отходов» и федеральным классификационным каталогом отходов (ФККО) компьютеры и периферийное оборудование, использовавшиеся при разработке и эксплуатации системы и утратившие потребительские свойства, относятся к IV-ому классу опасности, ртутные и люминесцентные лампы, использовавшиеся для создания искусственного освещения, относятся к I-ому классу опасности, использованная бумага и канцелярия относятся к IV-ому и V-ому классам опасности [23]. Отходы должны быть пройдены этапы технологического цикла отходов, подлежащих ликвидации, и утилизированы в соответствии со своим классом опасности согласно ГОСТ Р 53692–2009 «Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов» [38]. В России на сегодня развито три направления утилизации электроники: рынок б/у электронных компонентов, вторичная переработка цветных металлов и вторичная переработка электроники, содержащей драгметаллы. Также для снижения влияния на экологию необходимо стремиться к снижению энергопотребления, то есть разрабатывать и внедрять системы с малым энергопотреблением и минимизировать простой

техники. Следует использовать современные ЭВМ с режимом пониженного потребления электроэнергии при длительном простое.

5.6 Безопасность в чрезвычайных ситуациях

При разработке и эксплуатации системы возможно возникновение следующих чрезвычайных ситуаций различного происхождения:

- природного (землетрясение, бури);
- техногенного (пожары, взрывы, внезапное обрушение зданий, аварии на коммунальных системах жизнеобеспечения, аварии на электростанциях).

Наиболее вероятной чрезвычайной ситуацией на рабочем месте в помещении, оборудованном электронно-вычислительными машинами, является возникновение пожара. Это может происходить по причине близкого расположения элементов электронных систем ПК относительно друг друга – из-за нагревания током изоляция проводов может оплавиться и оголить провода, что приведет к короткому замыканию и искрению. Неисправность аппаратуры, неправильное использование электрооборудования, незнание мер безопасности и их пренебрежение также могут привести к возникновению пожара. Для обеспечения пожарной безопасности необходимо устранить потенциальные причины возникновения пожара в электрооборудованиях – это предупреждение замыкания грамотными выбором, монтажом и эксплуатацией сетей, работа только с исправным оборудованием и электропроводкой. Помещение должно быть оснащено рабочими порошковыми или углекислотными огнетушителями. Недопустимо применение в качестве средств пожаротушения электроприборов воды или пены из-за опасности поражения электрическим током. Необходимо обеспечить возможность беспрепятственного движения людей по эвакуационным путям [24]. Перед работой необходимо провести для персонала инструктаж по технике безопасности. В случае возникновения пожара необходимо обратиться в пожарную службу. Если возпламенение небольшое и его по силам потушить имеющимися средствами пожаротушения, то это необходимо сделать. В ином случае следует приступить к эвакуации из здания в соответствии с планом эвакуации при пожарах и других ЧС.

Согласно СП 12.13130.2009 наиболее вероятные классы пожаров в помещениях с ПЭВМ – «А» и «Е». Могут гореть в основном твердые вещества, горение которых сопровождается тлением – класс А или возможны пожары, вызванные возгоранием электроустановок – класс Е. Согласно статье 123 ФЗ к первичным средствам пожаротушения в помещениях с ПЭВМ относятся различные углекислотные, аэрозольные, порошковые огнетушители, предназначенные для тушения загораний и пожаров в начальной стадии их развития. Кроме того, необходим инструктаж сотрудников по плану действий в случае возникновения чрезвычайной ситуации.

5.7 Выводы по разделу

В ходе написания раздела были рассмотрены правовые, организационные и производственные вопросы обеспечения безопасности человека при работе с системой. Были проанализированы вредные и опасные факторы, предложены мероприятия по их предотвращению. Также были рассмотрены вопросы негативного влияния на окружающую среду, были предложены меры по обеспечению экологической безопасности. Было проанализировано возможное возникновение ЧС на рабочем месте, предложены рекомендации по профилактике возникновения наиболее вероятной ЧС – пожара.

Таблица 5.4 – Соответствие фактических значений потенциально возможных факторов нормативным значениям

Производственный фактор	Показатель	Реальное значение	Нормативное значение
Умственное перенапряжение, в том числе вызванное информационной нагрузкой	Класс условий труда	3.1 – напряженный труд 1 степени	2 – допустимый
Отсутствие или недостаток необходимого естественного освещения	КЕО ен, %, при боковом освещении	0,6	От 0,5
Производственные факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под	Напряжение (U) и сила тока (I) при прикосновении	При целости оборудования удары током невозможны	U не более 2 В, I не более 0,3 мА

действие работающий	которого	попадает			
------------------------	----------	----------	--	--	--

- Согласно «Правилам устройства электроустановок» рабочее помещение является нормальным, без повышенной опасности.

- Согласно «Правилам по охране труда при эксплуатации электроустановок» персонал относится к I группе по электробезопасности.

- Согласно СанПиН 1.2.3685–21 "Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания", тяжесть труда принадлежит к категории I.

- Согласно СП 12.13130.2009 «Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности», взрывопожарная и пожарная опасность к категории В4 – пожароопасность.

- Согласно критериям отнесения объектов, оказывающих негативное воздействие на окружающую среду, к объектам I, II, III и IV категорий, объект относится к IV категории.

ЗАКЛЮЧЕНИЕ

В процессе выполнения работы была реализована система, позволяющая сравнивать файлы студенческих работ, расположенные на локальном диске, и определять подобные из них указанием процента заимствования.

Были проанализированы существующие алгоритмы и получены обширные знания об алгоритмах обнаружения подобия. Проведено изучение существующих систем антиплагиата и дальнейшее их сравнение по ряду параметров, что также дало глубокое понимание процесса. Весь необходимый функционал системы реализован с высоким уровнем эффективности на текущем этапе. Так как большинство отчетов студентов оформляется в формате *.doc, то была реализована возможность открытия и загрузки документов Word. С помощью языка Dart эта проблема была решена после глубоких исследований.

В ходе выполнения ВКР был подробно изучен алгоритм шинглов, лежащий в основе анализа подобия файлов системы. Алгоритм шинглов имеет большую пространственную и вычислительную сложность, а также высокую точность обнаружения подобия. Однако он требует некоторой оптимизации, потому что, когда количество файлов чрезмерно велико, системные накладные расходы велики, если обрабатываются все подобия, включая ресурсы памяти и ЦП. Он оптимален для приложений с небольшими объемами данных и высокими требованиями к точности и должен быть достаточным для нашего использования.

В систему был успешно добавлен способ кластеризации результатов и включение одновременного много файлового сравнения. Функция фильтрации результатов на основе процентного сходства сработала эффективно.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Gaël Thomas (2022) What is Flutter and why you should learn it [Электронный ресурс]. – // URL: <https://herewecode.io/blog/what-is-flutter-and-why-you-should-learn-it>. (дата обращения 20.04.2022)
2. Yahia Jazyah. (2018) Open Learning, the Issue of Plagiarism - Efficient Algorithm. International Journal of Computers, 3, 105-111.
3. H. Maurer, F. Kappe e B. Zaka, Plagiarism - A Survey, Journal of Universal Computer Science, 2006, 12(8), pp. 1050-1084. – // URL: http://jucs.org/jucs_12_8/plagiarism_a_survey/jucs_12_08_1050_1084_maurer.pdf
4. A. R. Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment", Proc. 4th Int. Conf. Cyber IT Service Manage., pp. 1-6, Apr. 2016.
5. Shofi Ullah, Sagar Hossain, K. M. Azharul Hasan, "Opinion Summarization of Bangla Texts using Cosine Simillarity Based Graph Ranking and Relevance Based Approach", 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), pp.1-6, 2019.
6. Pongsakorn Saipech, Pusadee Seresangtakul, "Automatic Thai Subjective Examination using Cosine Similarity", 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), pp.214-218, 2018.
7. Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D. Text Classification Algorithms: A Survey. Information. 2019; 10(4):150. <https://doi.org/10.3390/info10040150>
8. A Gentle Introduction to the Bag-of-Words Model by Jason Brownlee on October 9, 2017 in Deep Learning for Natural Language Processing <https://machinelearningmastery.com/gentle-introduction-bag-words-model/>
9. Prasetya, Didik & Wibawa, Aji & Hirashima, Tsukasa. (2018). The performance of text similarity algorithms. International Journal of Advances in Intelligent Informatics. 4. 10.26555/ijain.v4i1.152.

10. "CiteSeerX (Pennsylvania State University)". – // URL: <https://citeseerx.ist.psu.edu>. [Электронный ресурс]. (Дата обращения 20.11.2021)
11. Bretag, T., & Mahmud, S. (2009). A model for determining student plagiarism: Electronic detection and academic judgement. *Journal of University Teaching & Learning Practice*, 6(1). – // URL: <http://ro.uow.edu.au/jutlp/vol6/iss1/6>
12. Foltýnek, Tomáš; Meuschke, Norman; Gipp, Bela (16 October 2019). "Academic Plagiarism Detection: A Systematic Literature Review". *ACM Computing Surveys*. 52 (6): 1–42. doi:10.1145/3345317.
13. Stein, Benno; Lipka, Nedim; Prettenhofer, Peter (2011), "Intrinsic Plagiarism Analysis" (PDF), *Language Resources and Evaluation*, 45 (1): 63–82, doi:10.1007/s10579-010-9115-y, ISSN 1574-020X, S2CID 13426762, archived from the original (PDF) on 2 April 2012, (дата обращения: 15.12.2021).
14. Методы выявления плагиата [Электронный ресурс]. – // URL: [https://commons.wikimedia.org/wiki/File:Методы выявления плагиата.png](https://commons.wikimedia.org/wiki/File:Методы_выявления_плагиата.png) (дата обращения: 15.11.2020).
15. What is Plagiarism? [Электронный ресурс]. – // URL: <https://www.plagiarism.org/article/what-is-plagiarism> (дата обращения: 4.01.2022).
16. How the shingle method works when checking text for plagiarism. Shingle - mysterious and incomprehensible Principle of the shingle method scientific articles [Электронный ресурс]. – // <https://abramov-2110.ru/en/kak-rabotaet-metod-shinglov-pri-proverke-teksta-na-plagiat-shingl/> (дата обращения: 16.12.2021).
17. Cosine Similarity – Understanding the math and how it works (with python codes). Prabhakaran S. [Электронный ресурс]. – // URL: <https://www.machinelearningplus.com/nlp/cosine-similarity> (дата обращения: 20.12.2021).
18. Трудовой кодекс [Электронный ресурс] / Справочная правовая система "Кодекс" URL: <https://docs.cntd.ru/document/901807664> (дата обращения: 11.05.2022).
19. ГОСТ 12.4.011–89 Система стандартов безопасности труда (ССБТ). Средства защиты работающих. Общие требования и классификация

[Электронный ресурс] / URL: <https://docs.cntd.ru/document/1200000277>, свободный. – Яз. рус. Дата обращения: 11.05.2022.

20. ГОСТ Р 12.1.019–2017 Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты [Электронный ресурс] / URL: <https://docs.cntd.ru/document/1200161238>, свободный. – Яз. рус. Дата обращения: 11.05.2022.

21. СП 52.13330.2016 Естественное и искусственное освещение [Электронный ресурс] / Электронный фонд правовой и нормативно-технической информации. URL: <https://docs.cntd.ru/document/456054197>, свободный. – Яз. рус. Дата обращения: 11.05.2022.

22. ГОСТ 12.1.038–82 Система стандартов безопасности труда. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов [Электронный ресурс] / Электронный фонд правовой и нормативно-технической информации. URL: <https://docs.cntd.ru/document/5200313>, свободный. – Яз. рус. Дата обращения: 11.05.2022.

23. Приказ Минприроды России от 30.09.2011 N 792 «Об утверждении Порядка ведения государственного кадастра отходов» токов [Электронный ресурс] / Электронный фонд правовой и нормативно-технической информации. URL: <https://docs.cntd.ru/document/902305590>, свободный. – Яз. рус. Дата обращения: 12.05.2022.

24. ГОСТ 12.1.004–91 Пожарная безопасность. Общие требования [Электронный ресурс] / Электронный фонд правовой и нормативно технической информации. URL: <https://docs.cntd.ru/document/9051953>, свободный. – Яз. рус. Дата обращения: 13.05.2022.

Приложение А

Таблица 2 – Результаты сравнения приложений

Файл	Plagiarism CheckerX, %	Plagiarism, %	Разница,%
1. Remizov	29	26.75	< 5
2. Брехова	30	26.22	< 5
3. Калипин	31	23.85	> 5
4. Молдабеков	30	24.80	< 5
5. Масслеников	26	23.22	< 5
6. Bulko	24	21.48	< 5
7. Devin	24	23.06	< 5
8. Domshoev	41	36.86	< 5
9. Банников	26	23.22	< 5
10. Бутырина	32	28.28	< 5