

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

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Непрерывная идентификация пользователей на основе динамических характеристик клавиатурного почерка и методов Data Mining.

УДК 004.93'1:004.353.4

Студент

Группа	ФИО	Подпись	Дата
8И6А	Затеев Роман Павлович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	К.Т.Н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Подопригора И.В.	К. Э. Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД	Немцова О.А.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ООП	Цапко И.В.	К.Т.Н.		
Руководитель ОИТ ИШИТР	Шерстнев В.С.	К.Т.Н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результата	Результат обучения (Выпускник должен быть готов)
Профессиональные компетенции	
P1	Применять базовые и специальные естественнонаучные и математические знания для комплексной инженерной деятельности по созданию, внедрению и эксплуатации геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием геоинформационных систем и технологий, информационных систем в бизнесе, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Выполнять комплексные инженерные проекты по созданию информационных систем и технологий, а также средств их реализации (информационных, методических, математических, алгоритмических, технических и программных).
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P6	Внедрять, эксплуатировать и обслуживать современные геоинформационные системы и технологии, информационные системы и технологии в бизнесе, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
Универсальные компетенции	
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Осуществлять коммуникации в профессиональной среде и в обществе в целом. Владеть иностранным языком (углубленный английский язык), позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена команды, состоящей из специалистов различных направлений и квалификаций.
P10	Демонстрировать личную ответственность за результаты работы и готовность следовать профессиональной этике и нормам ведения комплексной и инженерной деятельности.
P11	Демонстрировать знания правовых, социальных, экологических и культурных аспектов комплексной инженерной деятельности, а также готовность достижению должного уровня физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности.

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

УТВЕРЖДАЮ:
 Руководитель ООП
 _____ Кочегурова Е.А.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

бакалаврской работы

Студенту:

Группа	ФИО
8И6А	Затееву Роману Павловичу

Тема работы:

Непрерывная идентификация пользователей на основе динамических характеристик клавиатурного почерка и методов Data Mining.	
Утверждена приказом директора (дата, номер)	28.02.2020 г. № 59-66/с

Срок сдачи студентом выполненной работы:	06.06.2020
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	<i>Работа направлена на создание алгоритмического и программного обеспечения системы распознавания клавиатурной динамики пользователей информационной системы</i>
Перечень подлежащих исследованию, проектированию и разработке вопросов	<ul style="list-style-type: none"> – Анализ предметной области, – изучение динамических характеристик клавиатурного почерка, – обзор методов идентификации пользователей по клавиатурному почерку, – сбор исходных данных, – разработка алгоритма идентификации, – разработка программного приложения, – тестирование и анализ результатов, – финансовый менеджмент, ресурсоэффективность и ресурсосбережение, – социальная ответственность
Перечень графического материала	<i>Презентация в формате *.pptx</i>

Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Финансовый менеджмент	Подопригора И.В.
Социальная ответственность	Немцова О.А.
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Все разделы должны быть написаны на русском языке.	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	27.01.2020
---	------------

Задание выдал руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	К.Т.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И6А	Затеев Р.П.		

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: Информационные системы и технологии

Отделение школы (НОЦ): Отделение информационных технологий

Период выполнения: весенний семестр 2019 /2020 учебного года

Форма представления работы:

бакалаврская работа

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	06.06.2020
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
16.06.20	<i>Основная часть</i>	75
18.04.20	<i>Финансовый менеджмент</i>	15
07.04.20	<i>Социальная ответственность</i>	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	к.т.н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Цапко И.В.	к.т.н.		

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

Студенту:

Группа	ФИО
8И6А	Затееву Роману Павловичу

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
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>	<i>Оклад инженера – 21760 руб. Оклад руководителя – 33664 руб.</i>
2. <i>Нормы и нормативы расходования ресурсов</i>	<i>Премимальный коэффициент 30%; Коэффициент доплат и надбавок 20%; Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%.</i>
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	<i>Коэффициент отчислений на уплату во внебюджетные фонды 30 %</i>

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	<i>SWOT-анализ; Сравнительная характеристика вариантов исполнения проекта.</i>
2. <i>Планирование и формирование бюджета научных исследований</i>	<i>Формирование плана и графика разработки: - определение структуры работ, - определение трудоемкости работ, - создание диаграммы Гантта. Формирование бюджета затрат на разработку: - материальные затраты, - заработная плата (основная и дополнительная), - социальные отчисления, накладные расходы.</i>
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности разработки</i>	<i>Определение потенциального эффекта исследования</i>

Перечень графического материала (с точным указанием обязательных чертежей):

1. *Оценка конкурентоспособности технических решений*
2. *Матрица SWOT*
3. *Альтернативы проведения НИ*
4. *График проведения и бюджет НИ*
5. *Оценка ресурсной, финансовой и экономической эффективности НИ*

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Подопригора И.В.	К. Э. Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И6А	Затеев Р.П.		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8И6А	Затееву Роману Павловичу

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

Исходные данные к разделу «Социальная ответственность»:

<p>1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения</p>	<p>Объектом исследования является разрабатываемая система распознавания пользователя по клавиатурному почерку. Система включает в себя методику анализа клавиатурного почерка, алгоритм распознавания и обучения, а также программное обеспечение.</p> <p>Использование системы происходит в офисных помещениях и требует работы с компьютерами, которые являются источниками вредных излучений и могут оказывать негативное влияние на здоровье и жизнедеятельность человека.</p>
---	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Правовые и организационные вопросы обеспечения безопасности:</p> <ul style="list-style-type: none"> - специальные правовые нормы трудового законодательства; - организационные мероприятия при компоновке рабочей зоны. 	<p>Используемая нормативно-техническая документация:</p> <ul style="list-style-type: none"> - ГОСТ 12.2.032-78 - СанПиН 2.2.4.548-96 - ГОСТ 12.2.032-78
<p>2. Производственная безопасность:</p> <p>2.1 Анализ выявленных вредных и опасных факторов</p> <p>2.2 Обоснования мероприятий по снижению воздействия</p>	<p>Анализ выявленных вредных факторов:</p> <ul style="list-style-type: none"> - Недостаточная освещенность рабочей зоны - Отклонение параметров микроклимата - Превышение уровня шума - Психофизические факторы - Напряженность магнитного поля - Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека - Повышенный уровень статического электричества
<p>3. Экологическая безопасность</p>	<p>В процессе разработки и эксплуатации системы идентификации возможно образование следующих видов отходов:</p> <ul style="list-style-type: none"> • Твердые отходы: бумага, канцелярские принадлежности, комплектующие; • Жидкие отходы: сточные воды; • Люминесцентные лампы.
<p>4. Безопасность в чрезвычайных ситуациях</p>	<p>Наиболее типичная чрезвычайная ситуация при работе в офисе – пожар. Превентивные меры включают инструктаж по пожарной безопасности, контроль состояния проводки и электрических приборов, своевременное профилактическое обслуживание.</p>

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД ШБИП	Немцова О.А.			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И6А	Затеев Р.П.		

РЕФЕРАТ

Выпускная квалификационная работа выполнена на 76 страницах, содержит 26 рисунков, 25 таблиц, 16 источников, 1 приложение.

Ключевые слова: клавиатурный почерк, биометрия, идентификация, информационная безопасность, Data Mining.

Объектом исследования является система распознавания клавиатурной динамики пользователей информационной системы.

Цель работы – создание алгоритмического и программного обеспечения системы распознавания клавиатурной динамики.

В процессе исследования проводились работы по изучению методов идентификации пользователей по клавиатурному почерку. В ходе работы были рассмотрены существующие подходы к решению задачи идентификации, а также предложены свои решения.

В результате исследования были изучены существующие, а также протестированы разработанные алгоритмы идентификации пользователя по клавиатурному почерку.

Область применения: идентификация пользователей по клавиатурному почерку, определение психофизиологического состояния пользователя корпоративной сети, скрытый мониторинг пользователей корпоративной сети с целью определения подмены оператора.

Оглавление

Введение	12
1 Вопросы и задачи идентификации пользователей по клавиатурному почерку	13
1.1 Современное состояние проблемы	13
1.2 Направления исследований, связанных с динамикой нажатия клавиш	14
1.3 Характеристики клавиатурного почерка	16
1.4 Алгоритмы аутентификации и идентификации	17
1.4.1 Оценка метрических расстояний	18
1.4.2 Метод К-ближайших соседей	20
1.5 Оценка эффективности идентификации	22
1.6 Задачи из реальной практики	24
1.7 Постановка задачи	25
2 Решение задачи идентификации пользователя	26
2.1 Решение задачи Data Mining	26
2.1.1 Сбор реальных данных	26
2.1.2 Моделирование сессий	27
2.2 Идентификация	28
2.3 Чувствительность методов	29
3 Программная реализация	30
3.1 Выбор инструментов для разработки	30
3.2 Моделирование	30
3.2.1 Описание приложения	30
3.2.2 Архитектура приложения	30

3.2.3	Описание интерфейса.....	32
3.2.4	Анализ результатов	36
3.3	Распознавание	36
3.3.1	Архитектура приложения.....	36
3.3.2	Описание интерфейса.....	38
3.3.3	Анализ результатов	42
4	Финансовый менеджмент.....	44
4.1	Предпроектный анализ.....	44
4.1.1	Потенциальные потребители результатов исследования	44
4.1.2	Анализ конкурентных технических решений	45
4.1.3	SWOT-анализ.....	46
4.2	Планирование научно-исследовательских работ.....	49
4.2.1	Структура работ в рамках научного исследования	49
4.2.2	Определение трудоемкости выполнения работ	50
4.2.3	Разработка графика проведения научного исследования.....	51
4.2.4	Бюджет научно-технического исследования	54
4.3	Вывод по разделу.....	57
5	Социальная ответственность.....	58
5.1	Правовые и организационные вопросы обеспечения безопасности	58
5.1.1	Правовые и организационные мероприятия по обеспечению безопасности.....	58
5.1.2	Специальные правовые нормы трудового законодательства	59
5.1.3	Организационные мероприятия при компоновке рабочей зоны	60

5.2	Производственная безопасность.....	60
5.2.1	Недостаточная освещенность рабочей зоны.....	61
5.2.2	Превышение уровня шума.....	62
5.2.3	Отклонение параметров микроклимата.....	62
5.2.4	Нервно-психические нагрузки.....	63
5.2.5	Перенапряжение зрительных анализаторов.....	64
5.3	Экологическая безопасность.....	68
5.4	Безопасность в чрезвычайных ситуациях.....	69
5.5	Вывод по разделу.....	71
	Заключение.....	72
	Conclusion.....	73
	Список литературы.....	73
	Приложение А.....	76

Введение

В связи с бурным развитием цифровых технологий становится всё более актуальной защита информации от несанкционированного доступа.

Постоянно растущая потребность в более безопасном контроле доступа к современным приложениям, традиционные методы, такие как ПИН-коды, токены или пароли, не справляются с поставленными задачами, так как они могут быть утеряны или украдены, что ставит под угрозу безопасность системы.

Одним из возможных подходов к решению данной задачи является использование биометрических данных человека, таких как отпечаток пальца, голос, изображение лица или радужная оболочка глаза, которые являются уникальными характеристиками человека и представляют собой отличное решение для повышения безопасности системы.

Одним из типов биометрических данных является клавиатурный почерк, который может быть использован в качестве решения безопасного беспарольного доступа пользователя к компьютеру. Ритмы и другие индивидуальные приемы, которые использует каждый пользователь при печати, уникальны так же, как рукописный почерк и электронная подпись. Основным преимуществом использования клавиатурного почерка являются доступность и отсутствие необходимости в наличии дополнительного оборудования. А дополнение системы идентификации режимом скрытого мониторинга делает систему удобной и ненавязчивой для пользователя. Все необходимые данные могут быть собраны без участия пользователя и даже без его ведома. Вместе с тем повышаются требования к эффективности математического, алгоритмического и программного обеспечения.

1 Вопросы и задачи идентификации пользователей по клавиатурному почерку

1.1 Современное состояние проблемы

В последние годы клавиатурная динамика активно используется в системах, обеспечивающих информационную безопасность.

Использование клавиатурного почерка для решения задачи аутентификации обладает рядом преимуществ:

1. Простота реализации и внедрения. Реализация исключительно программная, ввод осуществляется со стандартного устройства ввода (клавиатуры), а значит использование не требует приобретения никакого дополнительного оборудования. Это самый дешевый способ аутентификации по биометрическим характеристикам субъекта доступа.

2. Пользователю нет необходимости осуществлять дополнительные действия, помимо печати.

3. Возможность скрытого мониторинга. Пользователь системы может быть не в курсе, что осуществляется скрытый анализ его клавиатурного почерка [1].

За первое полугодие 2019 года Аналитическим центром InfoWatch зарегистрировано 1276 случаев утечки конфиденциальной информации. Это на 22% больше, чем за аналогичный период 2018 года (1039 утечек). На графике (рисунок 1.1) представлена динамика количества утечек, начиная с первого полугодия 2006 г.

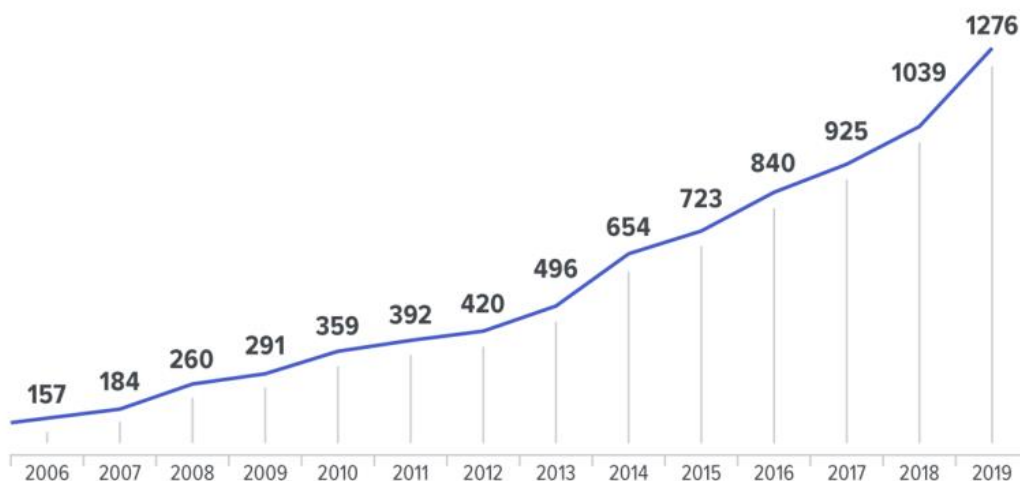


Рисунок 1.1 – Динамика утечек конфиденциальной информации

Согласно данным Аналитического центра, совокупная доля утечек персональных и платежных данных составила 85,1%. При этом на персональные данные пришлось 74,3% утечек (рост на 5,3 п. п. по сравнению с первым полугодием 2018 года). По сравнению с 2018 годом, снизилась доля утечек платежной информации, составив 10,8%. Заметно подросла доля утечек информации, составляющей коммерческую тайну. Доля утечек государственной тайны в мировом масштабе незначительно сократилась (рисунок 1.2) [2].

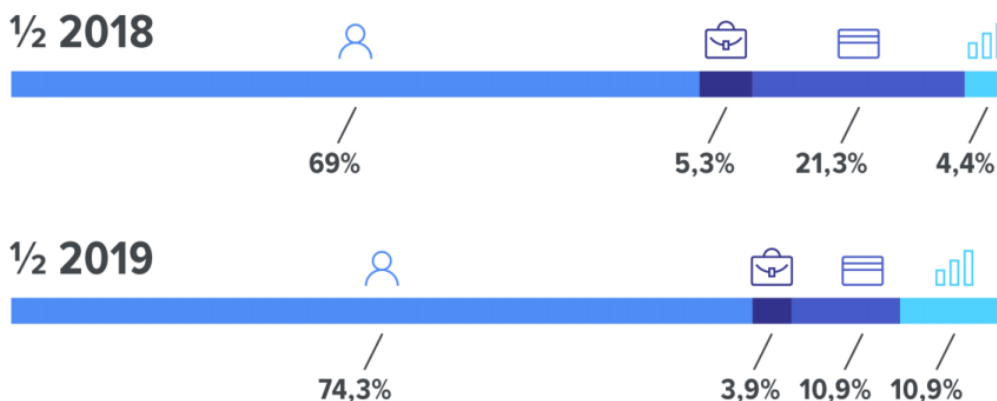


Рисунок 1.2 – Распределение утечек по типам данных

1.2 Направления исследований, связанных с динамикой нажатия клавиш

Существует два основных направления исследований, связанных с динамикой нажатия клавиш. С одной стороны, есть исследования, рассматривающие фиксированный и, как правило, короткий текст, в то время как, с другой стороны, есть исследования, сфокусированные на свободном тексте. Каждый способ имеет свои преимущества и недостатки.

Основные характеристики каждого из направлений:

- Фиксированный текст.

С помощью этой методики пользователю предлагается несколько раз ввести predetermined текст. Этот текст всегда одинаков (например, может быть пароль, логин или любая другая фиксированная строка). Этот вход при использовании статистических методов или методов машинного обучения будет использоваться для построения модели или для обучения системы маркировке, какие образцы принадлежат пользователю, а какие нет. Позже, когда

пользователи должны быть идентифицированы, они снова введут тот же текст, и этот новый образец будет сравнен с ранее сохраненным шаблоном или введен в алгоритм машинного обучения.

- Динамический текст.

В этом случае пользователи вводят либо длинные фрагменты текста, имитирующие идею свободного текста, либо, в других случаях, вводят все, что хотят без ограничений. Используя данный подход, работа выбранного алгоритма заключается в извлечении соответствующих функций и построении модели для каждого пользователя индивидуально. Когда позже пользователи должны быть идентифицированы, они могут ввести тот же либо совершенно другой текст. Выбранный алгоритм должен определить, являются ли новые образцы действительными или нет, то есть принадлежат ли они пользователю, претендующему на авторство.

Данные подходы имеют свои преимущества и недостатки в зависимости от модели, в которой они применяются. Например, если пользователи проходят проверку подлинности с использованием динамики нажатия клавиш и подхода с фиксированным текстом, политики доступа пользователей могут применяться только на этапе проверки подлинности. Это означает, что пользователь мог аутентифицироваться правильно, но ничего нельзя сказать о пользователе, который действительно использует компьютер после завершения этой фазы аутентификации. С другой стороны, задача сбора информации о динамике нажатия клавиш только на этом начальном этапе намного проще, чем при использовании подхода с произвольным текстом. Свободный текст можно использовать для постоянного мониторинга пользователя с помощью системы и применения различных политик, если обнаружено изменение в динамике нажатия клавиш. Недостатком этого способа является то, что пользователь должен постоянно находиться под наблюдением, и это может потреблять системные ресурсы и, в то же время, нарушать конфиденциальность пользователя при выполнении определенных задач, связанных с личной или конфиденциальной информацией.

1.3 Характеристики клавиатурного почерка

В исследованиях выделяют следующие характеристики клавиатурного почерка:

- Время удержания клавиши (ВУК)
- Число ошибок при печати
- Паузы между нажатиями
- Скорость набора
- Наличие наложений клавиш
- Частота использования служебных клавиш

Большая часть исследователей КП анализируют в своих работах временные характеристики между двумя последовательными нажатиями клавиш, так называемые диграфы (Di-graph). Существует два основных диграфа: время ожидания (Dwell time DT, Hold time HD) и время задержки (Flight time FT или *keystroke latency*). Времена ожидания и задержки имеют наглядную геометрическую иллюстрацию. На рисунке 1.3 приведены основные характеристики клавиатурного почерка [3].

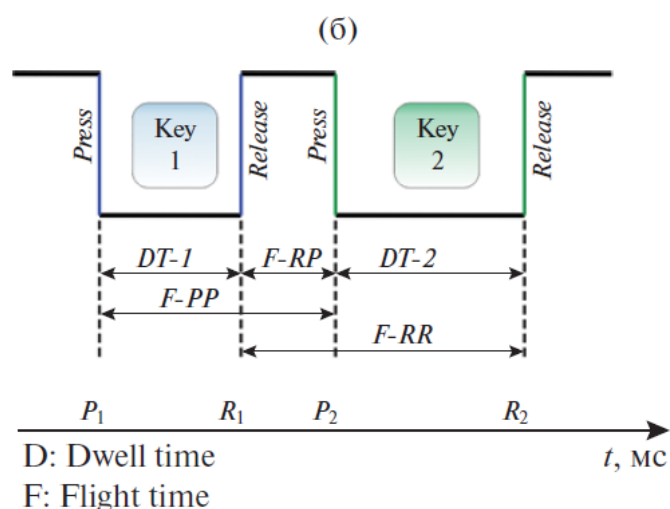


Рисунок 1.3 – Диграфы

Существует ряд ограничений по применению данного способа на практике. Применение способа идентификации по клавиатурному почерку целесообразно по отношению к пользователям с достаточно длительным опытом работы с компьютером и сформировавшимся почерком работы на клавиатуре [4].

Исходя из теории машинописи и делопроизводства можно определить время становления почерка работы с клавиатурой, при котором достигается необходимая вероятность идентификации пользователя: примерно 6 месяцев [5].

1.4 Алгоритмы аутентификации и идентификации

Алгоритмы статической и динамической аутентификации можно разделить на три основные группы с точки зрения распознавания образов:

- Оценка метрических расстояний.
- Статистические методы.
- Машинное обучение.

Относительная частота использования различных методов классификации показана на рисунке 1.4.

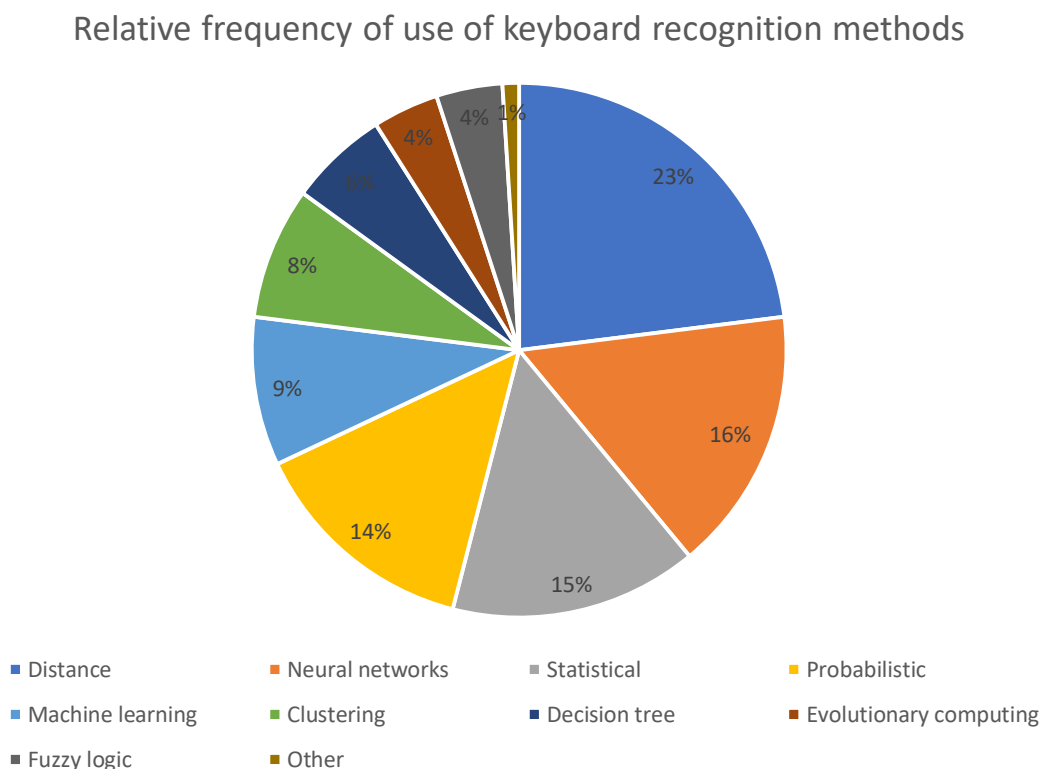


Рисунок 1.4 – Относительная частота использования методов распознавания клавиатурного почерка

Наибольшую группу методов составляют методы оценки близости между текущим и эталонным шаблоном пользователя путем вычисления некоторого метрического расстояния и сравнение его с пороговым значением. Исторически, статистические и вероятностные были первыми методами, использованными для

распознавания клавиатурного почерка. Как видно из диаграммы, данные методы не потеряли своей актуальности. Они используются в научных исследованиях и по сей день.

Следующий наиболее частый метод (16%) – это искусственные нейронные сети (ИНС). Они превратились в отдельную категорию среди методов машинного обучения и распознавания образов; однако эта группа распознавания клавиатурного почерка также включает в себя ряд известных алгоритмов, таких как нечеткая логика, деревья решений и генетические алгоритмы.

Кластерный анализ (8%) в распознавании клавиатурного почерка используется для объединения функций анализируемого шаблона с аналогичным профилем клавиатуры.

Разделение методов распознавания КП по категориям довольно интуитивно понятно. Одни и те же классификаторы могут основываться на разных методах обучения и метрических расстояниях.

1.4.1 Оценка метрических расстояний

Чтобы определить однотипность объектов нужно составить вектор характеристик для каждого объекта — как правило, это набор числовых значений.

В случае динамики нажатия клавиш в качестве характеристик могут выступать:

- Время удержания или длительность нажатия на клавишу
- Время между нажатиями клавиш
- Время между нажатиями на одноименные клавиши
- Процент ошибочных нажатий
- Скорость набора текста

После того, как определен вектор характеристик, можно провести нормализацию, чтобы все компоненты давали одинаковый вклад при расчете «расстояния». В процессе нормализации все значения приводятся к некоторому диапазону, например, $[-1, 1]$ или $[0, 1]$. Для каждой пары объектов измеряется «расстояние» между ними — степень похожести.

Существует множество методов оценки метрических расстояний:

1. Евклидово расстояние

Наиболее распространенная функция расстояния. Представляет собой геометрическим расстоянием в многомерном пространстве:

$$\rho(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1.1)$$

2. Квадрат евклидова расстояния

Применяется для придания большего веса более отдаленным друг от друга объектам. Это расстояние вычисляется следующим образом:

$$\rho(x, y) = \sum_i^n (x_i - y_i)^2 \quad (1.2)$$

3. Расстояние городских кварталов (манхэттенское расстояние)

Это расстояние является средним разностей по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако для этой меры влияние отдельных больших разностей (выбросов) уменьшается (т.к. они не возводятся в квадрат).
Формула для расчета манхэттенского расстояния:

$$\rho(x, y) = \sum_i^n |x_i - y_i| \quad (1.3)$$

4. Расстояние Чебышева

Это расстояние может оказаться полезным, когда нужно определить два объекта как «различные», если они различаются по какой-либо одной координате. Расстояние Чебышева вычисляется по формуле:

$$\rho(x, y) = \max(|x_i - y_i|) \quad (1.4)$$

5. Расстояние Махаланобиса

Расстояние Махаланобиса – это мера расстояния между векторами случайных величин, обобщающая понятие евклидова расстояния. С помощью расстояния Махаланобиса можно определять сходство неизвестной и известной

выборки. Оно отличается от расстояния Евклида тем, что учитывает корреляции между переменными и инвариантно к масштабу.

$$d(X, Y) = \sqrt{(X - Y) S^{-1} (X - Y)^T} \quad (1.5)$$

где S^{-1} – обратная корреляционная матрица;

X и Y – вектора значений.

Корреляционная матрица представляет собой квадратную матрицу размерности $M \times M$, где M – количество значений, симметричную относительно главной диагонали [6];

6. Степенное расстояние

Мера применяется в случае, когда необходимо увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие объекты сильно отличаются. Степенное расстояние вычисляется по следующей формуле:

$$\rho(x, y) = r \sqrt[r]{\sum_i^n (x_i - y_i)^p} \quad (1.6)$$

где r и p – параметры, определяемые пользователем. Параметр p ответственен за постепенное взвешивание разностей по отдельным координатам, параметр r ответственен за прогрессивное взвешивание больших расстояний между объектами. Если оба параметра – r и p – равны двум, то это расстояние совпадает с расстоянием Евклида [7].

Для повышения эффективности методов идентификации было принято решение использовать в вычислениях частотность букв русского языка, которая представлена в таблице А.1 в приложении А.

Зачастую, в сессиях реальных пользователей буквы ф, ъ, ё не встречались, либо встречались достаточно редко, поэтому данные буквы были изъяты из эксперимента.

1.4.2 Метод К-ближайших соседей

Метод К-ближайшего соседа – один из методов решения задачи классификации.

Предполагается, что уже имеется какое-то количество объектов с точной классификацией. То есть для каждого из них точно известно, к какому классу он принадлежит. Для решения задачи классификации нужно выработать правило, позволяющее отнести новый объект к одному из возможных классов. При этом все классы известны заранее.

В основе метода лежит следующее правило: объект считается принадлежащим тому классу, к которому относится большинство его ближайших соседей.

При использовании метода необходимо уметь определять, насколько объекты близки друг к другу. Это решается путем расчета расстояния между ними [8].

В данном исследовании расстояние было рассчитано между шаблоном пользователя и «новой» сессией. Расстояние рассчитывается исходя из количества ближайших букв.

Визуализация работы метода изображена на рисунках 1.5, 1.6, 1.7.

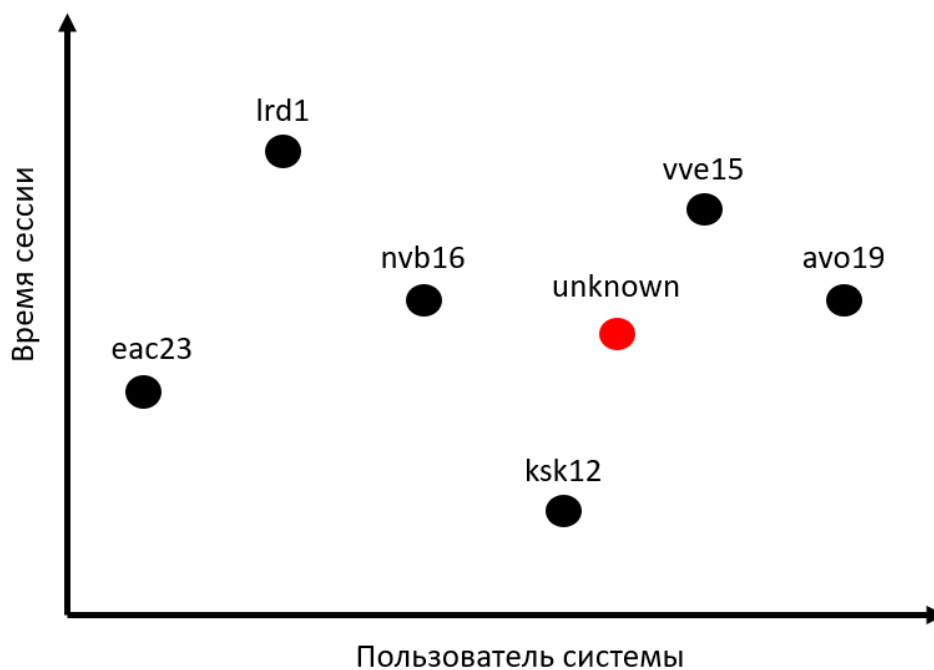


Рисунок 1.5 – Визуализация кластеров

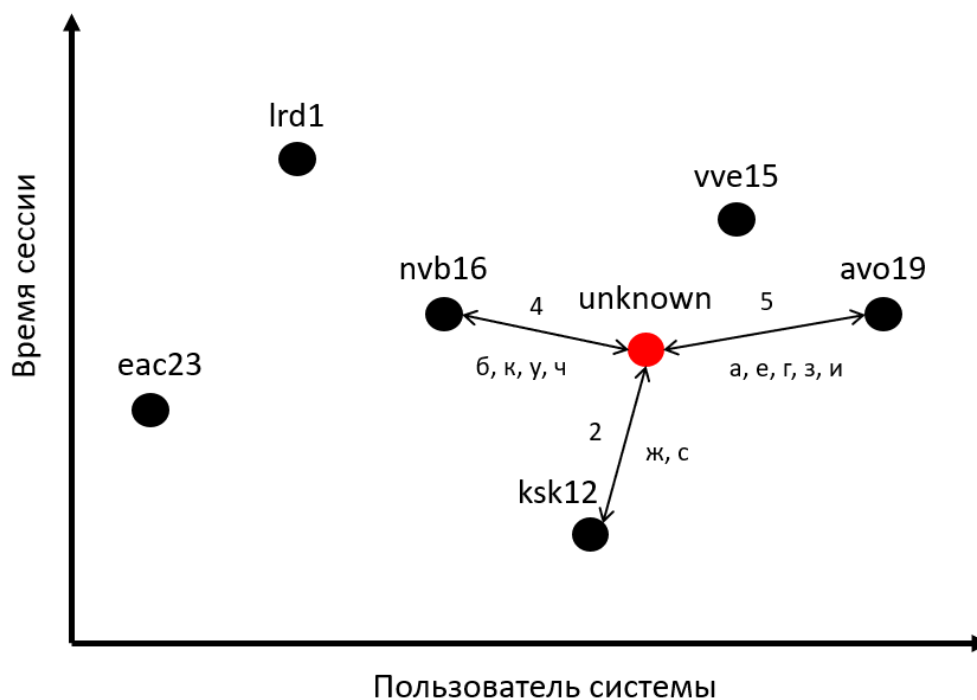


Рисунок 1.6 – Визуализация буквенных совпадений кластеров

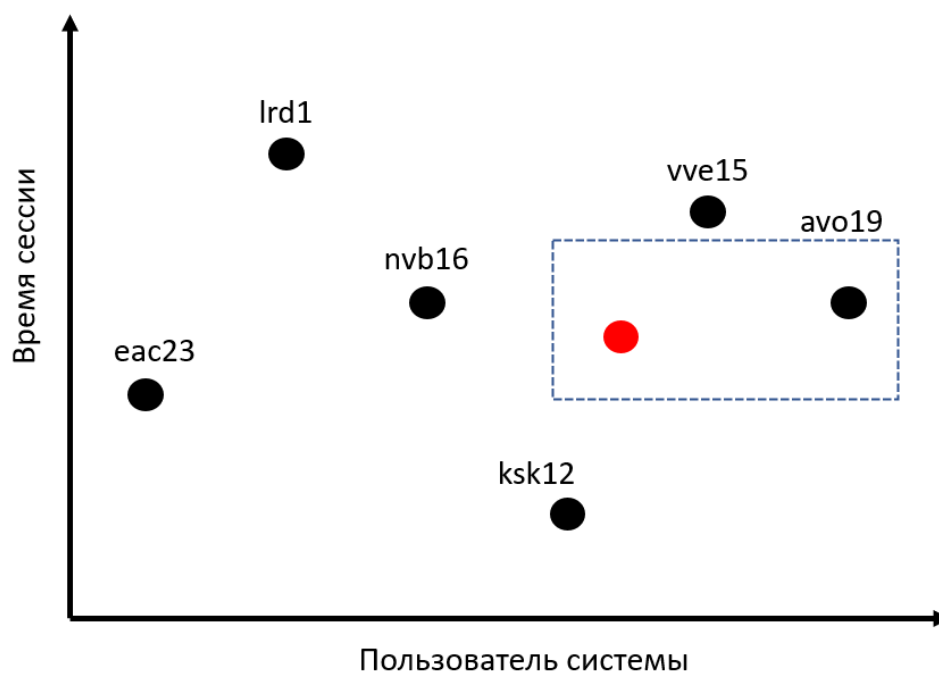


Рисунок 1.7 – Результат кластеризации

Таким образом, шаблон, набравший наибольшее количество буквенных совпадений, считается ближайшим. Далее, сессия идентифицируется логином ближайшего соседа.

1.5 Оценка эффективности идентификации

При парольной или статической идентификации проверяется эквивалентность вводимого текста эталонному. При динамической

идентификации оценивается вероятность ложного доступа незарегистрированного пользователя и вероятность отказа в доступе зарегистрированному пользователю.

False Rejection Rate (FRR) – оценка ложного отклонения, известна в статистической радиотехнике, как ошибка I рода. FRR определяет процент случаев отказа в доступе, когда законный пользователь ошибочно отклоняется. Вычисляется, как отношение ложно отклоненных пользователей к общему количеству пользователей системы.

$$FRR = \frac{\text{Количество ложных отказов}}{\text{Общее количество попыток}} \quad (1.7)$$

Более низкая FRR подразумевает меньшее отклонение пользователей и более легкий доступ для всех, в том числе и для незарегистрированных.

False Acceptance Rate (FAR) - ошибка ложного принятия. FAR или ошибка II рода определяет процент случаев принятия нелегальных пользователей. Вычисляется, как отношение ложно принятых неавторизованных пользователей к общему количеству пользователей системы [3].

$$FAR = \frac{\text{Количество ложных совпадений}}{\text{Общее количество попыток}} \quad (1.8)$$

Equal Error Rate – равная частота ошибок, предназначенная для определения общей точности системы распознавания. ERR представляет значения ошибки, когда FAR и FRR принимают равные значения. ERR не зависит от уровня чувствительности алгоритма.

Гипотетически ошибки FRR и FAR варьируются в зависимости от уровня чувствительности алгоритма, так называемого порогового значения: когда одна ошибка уменьшается, другая увеличивается (рисунок 1.8) [9].

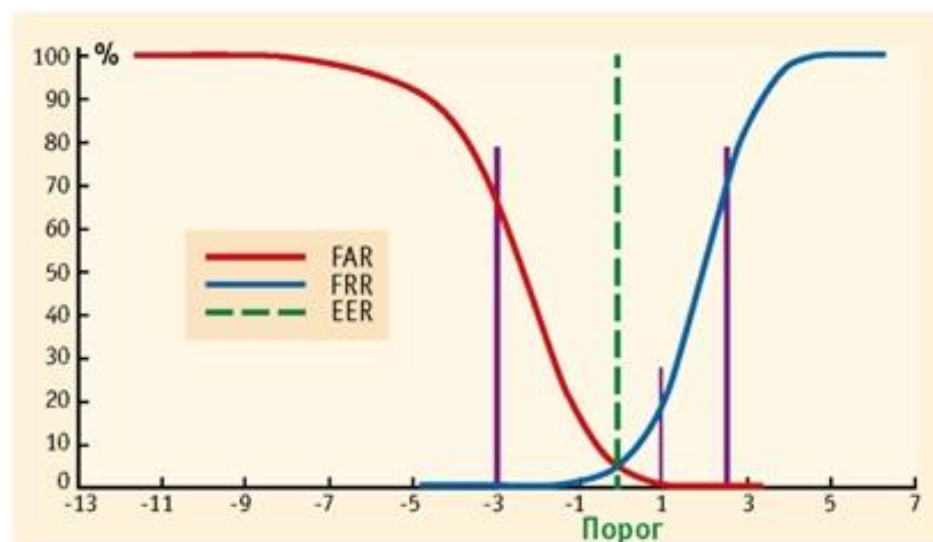


Рисунок 1.8 – Графики ошибок I и II рода

Если безопасность системы не имеет первостепенной важности, предпочтительнее более высокое значение FAR. Если разработчик хочет добиться более высокой степени защищенности системы, то стоит установить более высокое значение FRR.

Между FAR и FRR существует компромисс, который зависит от конкретной прикладной задачи. Добиться его можно, регулируя порог принятия решения.

1.6 Задачи из реальной практики

Постоянный скрытый мониторинг динамических характеристик клавиатурного почерка позволяет решить ряд прикладных задач в области защиты информации:

- Своевременное определение подмены пользователя. В случае отсутствия на рабочем месте аутентифицированного пользователя возможен доступ к конфиденциальной информации при незаблокированном компьютере.

- Идентификация психофизического состояния пользователя. При сильной усталости клавиатурный почерк человека меняется, становится менее ритмичным, увеличиваются паузы между нажатиями и количество ошибок ввода. Аналогичные проблемы возникают в состоянии алкогольного и наркотического опьянения. Присутствие на работе сотрудников в таком

состоянии контрпродуктивно, а в некоторых случаях сопряжено с риском для жизни и здоровья.

– Резервный метод аутентификации в случае взлома учётной записи, либо при утрате пароля или e-токена.

– Установление авторства документов, программного кода и т.д. в системах онлайн обучения, при проведении онлайн экзаменов и тестов. Это особенно актуально для дистанционного образования, для удаленного трудоустройства и формирования доступной инклюзивной среды.

1.7 Постановка задачи

Результатом ВКР будет являться программное приложение, предназначенное для непрерывной идентификации пользователей компьютерной сети, на основании динамических характеристик клавиатурного почерка.

2 Решение задачи идентификации пользователя

Идентификация – это процедура, в результате выполнения которой происходит распознавание субъекта по его идентификатору (логину, паролю и т.д.). Для выполнения процедуры идентификации в информационной системе субъекту предварительно должен быть назначен соответствующий идентификатор. То есть пользователь должен быть зарегистрирован в информационной системе [10].

В данном исследовании идентификация будет проводиться по шаблону клавиатурной динамики пользователя.

Решение задачи идентификации было разделено на 3 части:

1. Решение задачи Data Mining
2. Идентификация
3. Чувствительность методов

2.1 Решение задачи Data Mining

Основной задачей методов Data Mining является анализ больших объемов данных с целью извлечения ценной информации.

В ходе работы было принято решение разделить задачу Data Mining на следующие составляющие:

- Сбор реальных данных
- Моделирование сессий

Данное решение обосновано сложностью сбора реальных сессий пользователей из-за нескольких ограничений:

1. Невозможность получения прав администратора на ЭВМ в ТПУ для установки стороннего ПО.

2. Вопрос о защите персональных данных. Система отслеживает все нажатия пользователя на клавиатуру, вследствие чего ей могут быть записаны конфиденциальные данные пользователя, такие как логины, пароли и т.д.

2.1.1 Сбор реальных данных

Сбор данных о динамике нажатия клавиш – это начальный и самый важный этап реализации клавиатурной идентификации. Его эффективность

полностью определена программным обеспечением и не требует дополнительного оборудования. При нажатиях на клавиатуру операционная система фиксирует код ANSI, связанный с нажатой клавишей, а также время её нажатия и отпуска. Событие нажатие клавиши можно измерить с точностью до миллисекунды. На этом и основан сбор первичных данных по динамике нажатия клавиш пользователями [3].

Для сбора реальных статистик пользователями вводилось по 1000 символов. Для обучения системы, то есть составления шаблона на пользователя, необходимо провести минимум 10 сессий.

В результате данного этапа была собрана статистика по шести пользователям корпоративной системы.

2.1.2 Моделирование сессий

Моделирование сессий пользователей является одним из важнейших этапов исследовательской работы.

Полученные ранее сессии использовались для создания «шаблонов» пользователей, а также подсчета статистических показателей, таких как математическое ожидание, дисперсия и среднеквадратическое отклонение.

В данной работе моделирование необходимо, исходя из того, что исследование и оценка эффективности ранее рассмотренных алгоритмов невозможна на имеющемся объеме данных.

2.1.2.1 Преобразование Бокса-Мюллера

Для моделирования сессий пользователей использовалось преобразование Бокса-Мюллера.

Преобразование Бокса-Мюллера – метод моделирования стандартных нормально распределённых случайных величин. Данный метод был опубликован Джорджем Боксом и Мервином Мюллером в 1958 году. Метод, в отличие, от методов, основывающихся на центральной предельной теореме, является точным.

Пусть x и y — независимые случайные величины, равномерно распределённые на отрезке $[-1, 1]$. Вычислим $s = x^2 + y^2$. Если окажется, что

$s > 1$ или $s = 0$, то значения x и y следует «выбросить» и сгенерировать заново. Как только выполнится условие $0 < s \leq 1$, по формулам 2.1 и 2.2 следует рассчитать z_0 и z_1 , которые будут независимыми величинами, удовлетворяющими стандартному нормальному распределению [11].

$$z_0 = x \sqrt{\frac{-2 \ln s}{s}} \quad (2.1)$$

$$z_1 = y \sqrt{\frac{-2 \ln s}{s}} \quad (2.2)$$

После преобразования Бокса-Мюллера был осуществлен переход к общему нормальному распределению.

2.2 Идентификация

Существует множество методов и метрик, которые используются для решения задачи идентификации по клавиатурному почерку.

В данном исследовании были изучены следующие:

1. Эвклидово расстояние
2. Эвклидово расстояние + частотность букв русского языка
3. Расстояние городских кварталов
4. Расстояние городских кварталов + частотность букв русского языка
5. Метод ближайшего соседа
6. Пороговое значение

Идентификация пользователя ведется путем сравнения введенного образца клавиатурного почерка с клавиатурными шаблонами, хранящимися в базе данных на сервере.

Возможны несколько сценариев:

1. Сессия текущего пользователя не похожа ни на один из имеющихся шаблонов. В таком случае системой выдается ошибка аутентификации, и пользователь идентифицируется, как неопознанный.

2. Образец клавиатурного почерка пользователя похож на один или несколько шаблонов, хранящихся в базе данных. В случае нескольких похожих

шаблонов системой выдается идентификатор наиболее вероятного. Условие похожести разбирается в следующем подпункте. Если образец похож на один из шаблонов, хранящихся на сервере, система идентифицирует пользователя, а также добавляет текущую сессию в список его сеансов на сервер.

2.3 Чувствительность методов

В предыдущем пункте описывалось решение проблемы идентификации. Важным этапом является установка условия похожести. Оно заключается в том, что при использовании любых методов распознавания необходимо учитывать, удовлетворяет ли образец правилу трёх сигм (рисунок) нормальной выборки, согласно которой, 99,7% значений лежат на расстоянии не более трёх стандартных отклонений.

Правило трёх сигм (рисунок 2.1) утверждает, что вероятность того, что случайная величина отклонится от своего математического ожидания более чем на три среднеквадратических отклонения, практически равна нулю. Правило справедливо только для случайных величин, распределенных по нормальному закону.

Если образец не удовлетворяет данному условию, система должна идентифицировать пользователя как неопознанного.

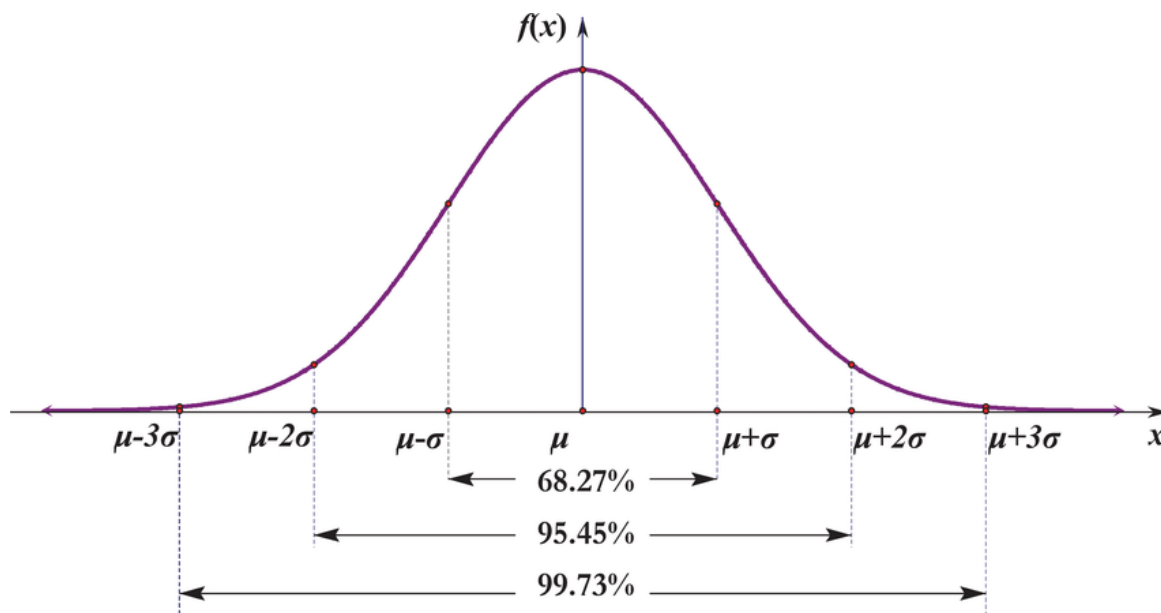


Рисунок 2.1 – Правило трёх сигм

3 Программная реализация

3.1 Выбор инструментов для разработки

Для разработки был выбран высокоуровневый язык программирования C#. C# был разработан для создания множества приложений, работающих в среде .NET Framework. Язык имеет строгую статическую типизацию, поддерживает полиморфизм, перегрузку операторов, указатели на функции-члены классов, атрибуты, события, свойства, исключения, комментарии в формате XML.

В качестве среды разработки использовалась Visual Studio 2019. Visual Studio представляет собой многофункциональную программу, которую может быть использована для различных аспектов разработки программного обеспечения. Помимо стандартного редактора и отладчика, Visual Studio включает в себя компиляторы, средства автозавершения кода, графические конструкторы и многие другие функции для упрощения процесса разработки.

3.2 Моделирование

3.2.1 Описание приложения

Приложение предназначено для моделирования сессий пользователей на основе реальных данных. В качестве метода моделирования используется преобразование Бокса-Мюллера.

3.2.2 Архитектура приложения

Диаграмма классов приложения изображена на рисунке 3.1.

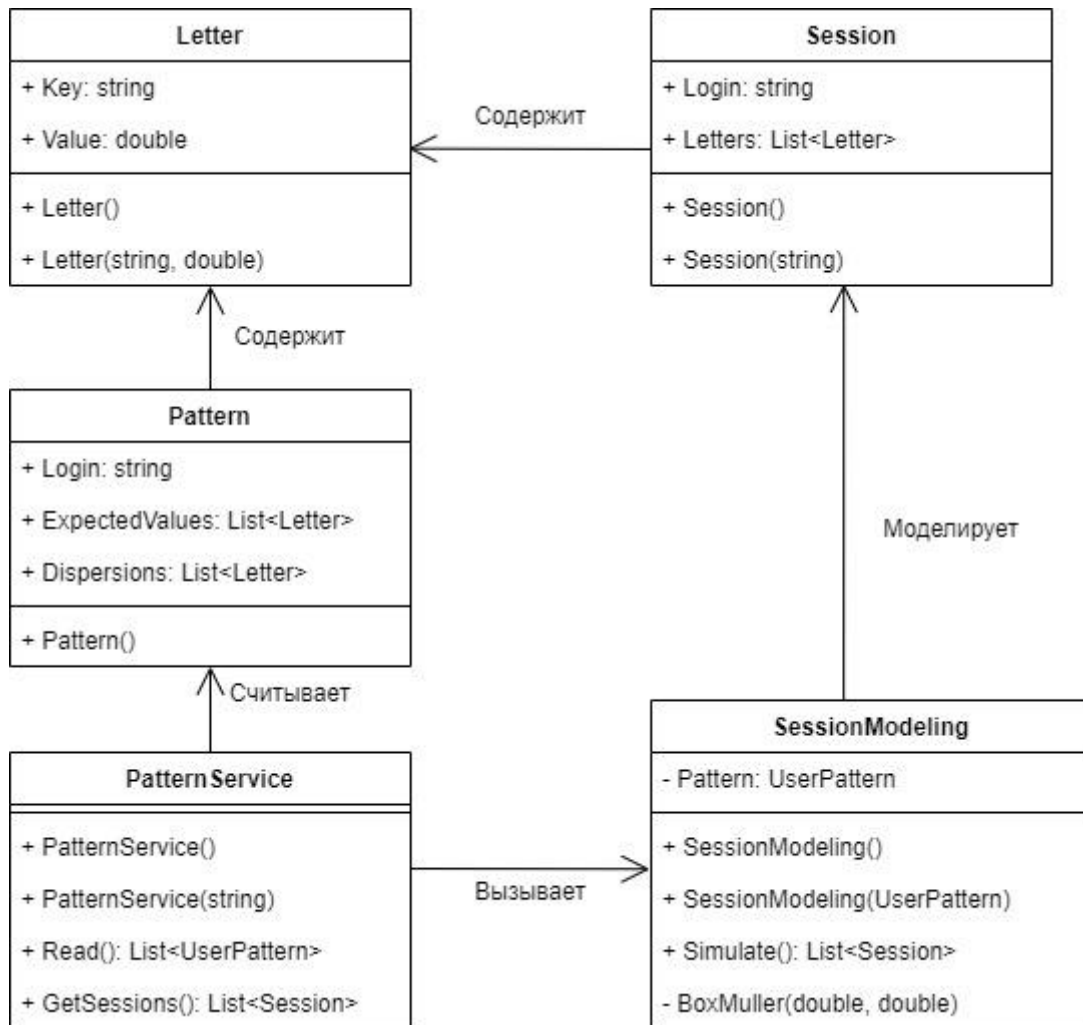


Рисунок 3.1 – UML-диаграмма классов приложения

Класс `PatternService` предназначен для получения шаблонов сессий из хранилища данных. Шаблоны сессий записываются в объекты класса `Pattern`.

Основным предназначением класса `SessionModeling` является моделирование сессий. Для моделирования в классе используется преобразование Бокса-Мюллера.

Смоделированные сессии записываются в объекты класса `Session`, которые, в последующем, добавляются в хранилище данных.

Класс `Letter` используется для описания буквы, нажатой пользователем, и временем удержания клавиши.

На рисунке 3.2 изображена диаграмма процесса «Моделирование сессий».

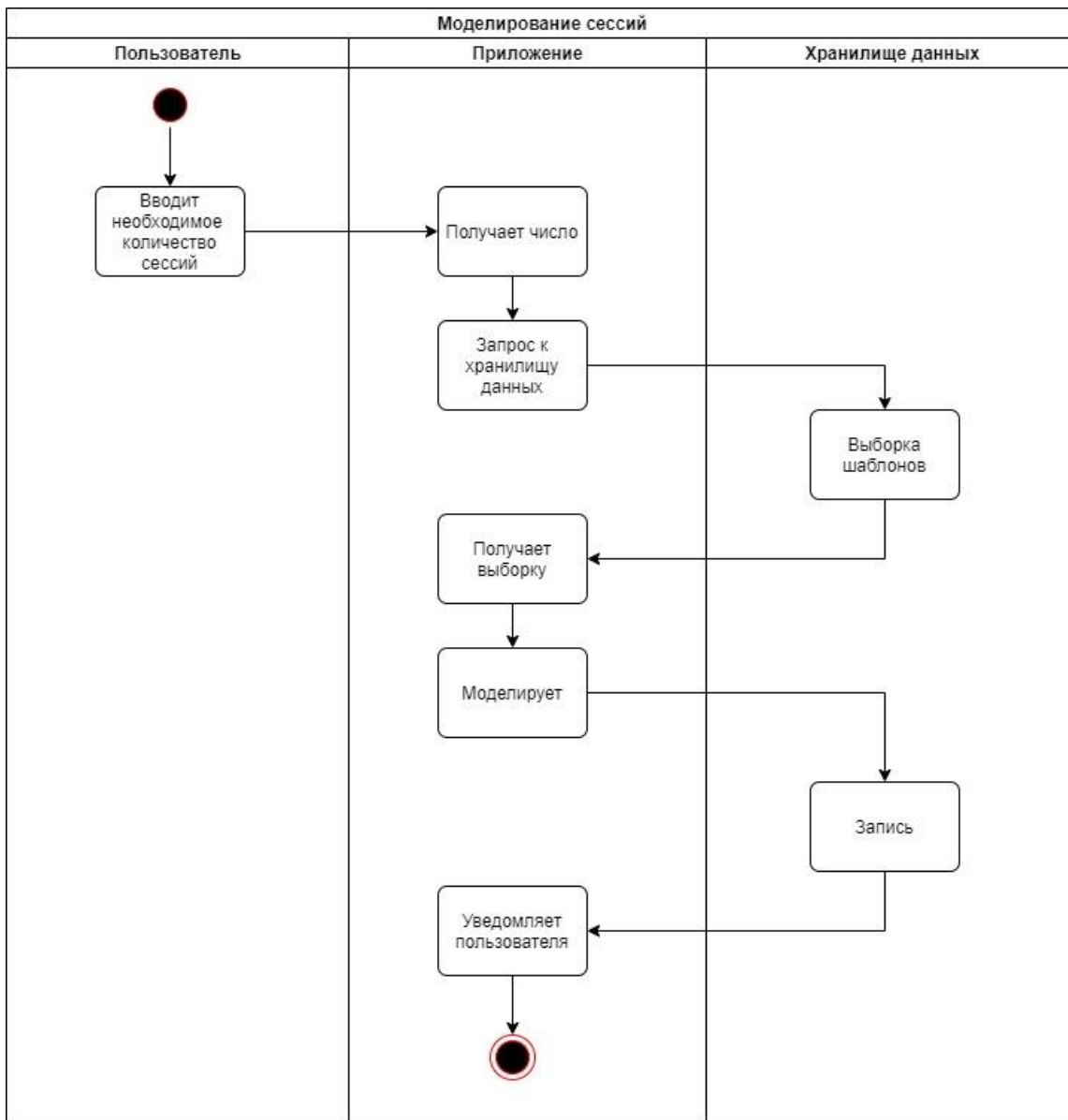


Рисунок 3.2 – Диаграмма процесса «Моделирование сессий»

3.2.3 Описание интерфейса

В состав приложения входят два окна (рисунок 3.3).

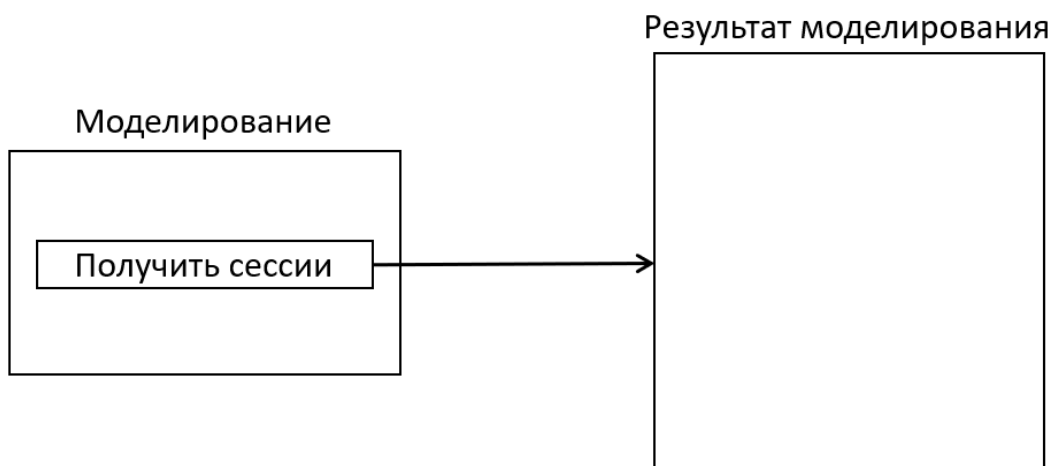


Рисунок 3.3 – Структура окон приложения

На рисунке 3.4 отображено основное окно приложения.

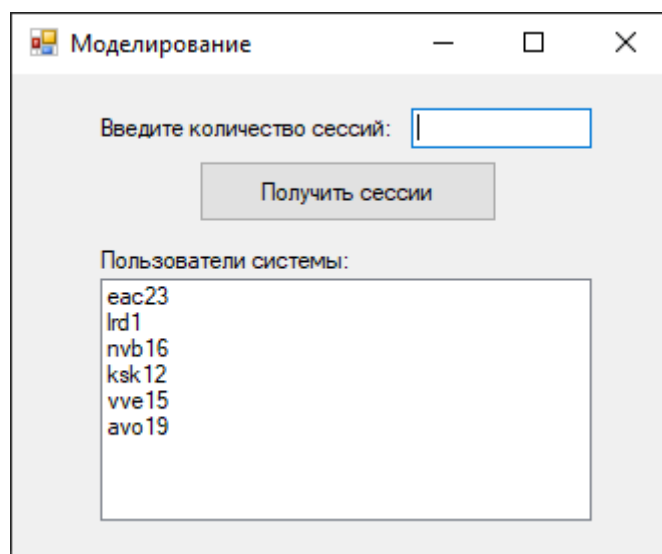


Рисунок 3.4 – Основное окно приложения

Окно приложения имеет в своем составе:

1. Поле «Количество сессий». Поле предназначено для ввода необходимого для моделирования количества сессий.
2. Кнопка «Получить сессии». Кнопка предназначена для запуска процесса моделирования
3. Список «Пользователи системы». В списке отображаются все пользователи, зарегистрированные в системе на момент запуска программы.

Для запуска процесса моделирования пользователю необходимо ввести необходимое количество сессий и выбрать пользователя системы, для которого будет осуществлено моделирование (рисунок 3.5).

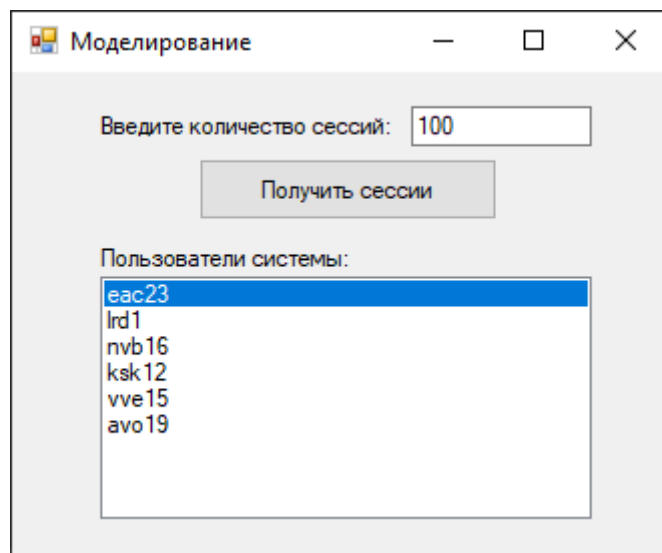


Рисунок 3.5 – Запуск процесса моделирования

После завершения процесса моделирования, появляется окно «Результат моделирования» (рисунок 3.6). В окне отображается результат моделирования сессий в формате JSON.

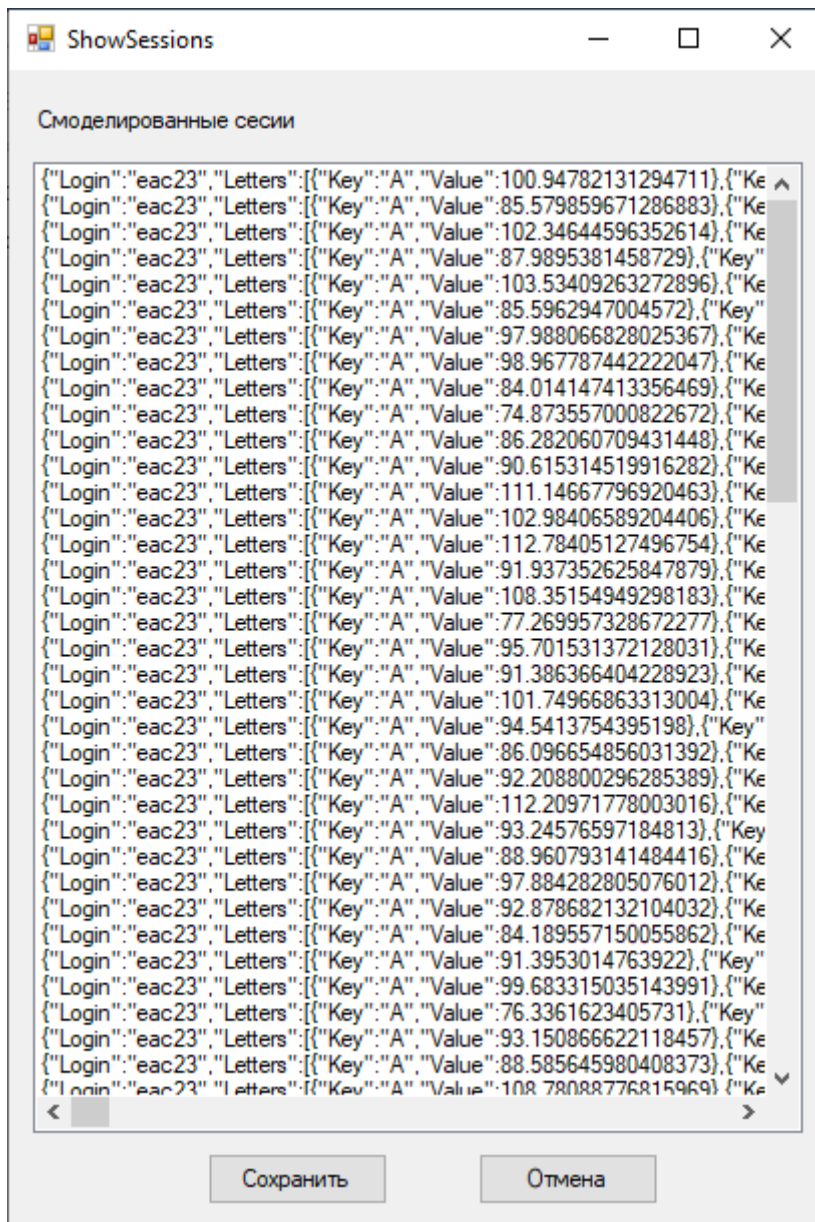


Рисунок 3.6 – Результат моделирования

Для сохранения смоделированных сессий пользователю необходимо нажать на кнопку «Сохранить», далее выбрать путь и имя файла.

После успешного сохранения отображается окно, изображенное на рисунке 3.7.

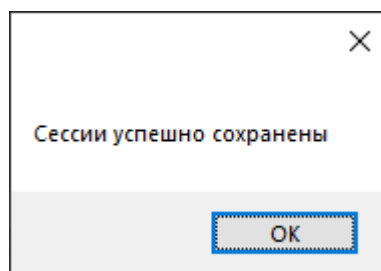


Рисунок 3.7 – Сообщение об успешном сохранении

3.2.4 Анализ результатов

Моделирование было выполнено из расчета 100 сессий по 1000 символов на одного пользователя.

График разброса времен нажатия на каждую из клавиш пользователя nvb16 для выборки из 10 сгенерированных сессий изображен на рисунке 3.8.

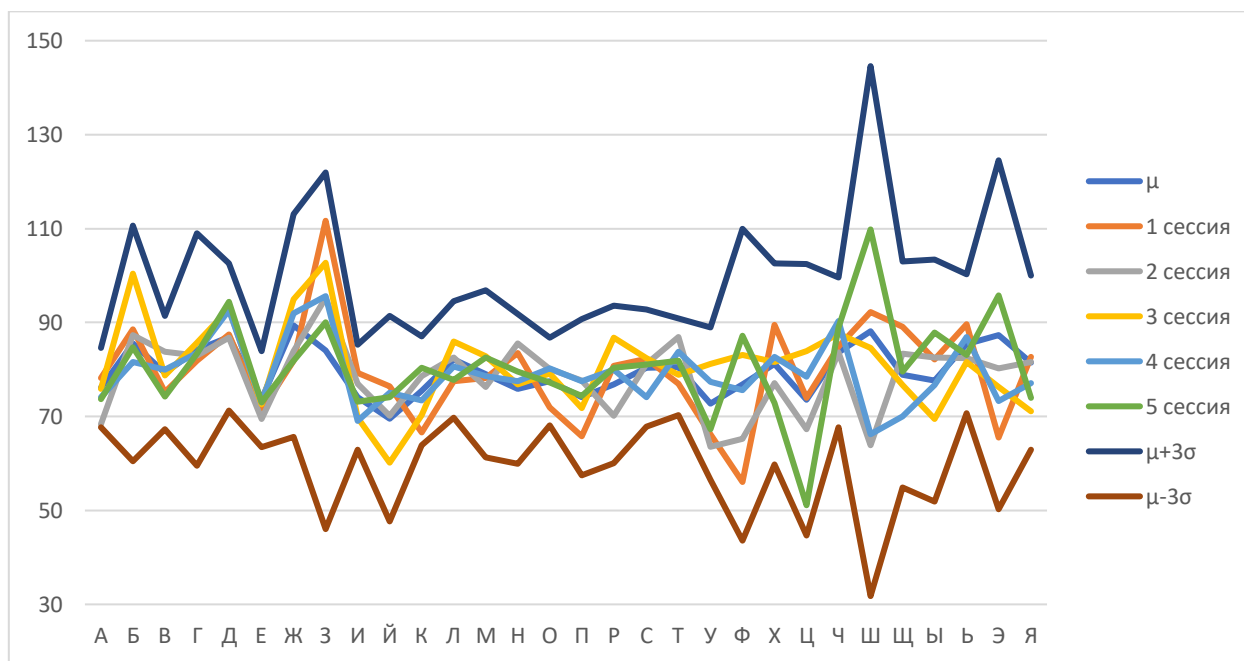


Рисунок 3.8 – График разброса значений сгенерированных сессий для пользователя nvb16

Как видно из рисунка, сгенерированные значения удовлетворяют правилу трёх сигм (рисунок 2.1).

3.3 Распознавание

Приложение предназначено для решения задачи идентификации пользователя в информационной системе. В приложении используются методы динамической идентификации.

3.3.1 Архитектура приложения

Диаграмма классов приложения изображена на рисунке 3.9.

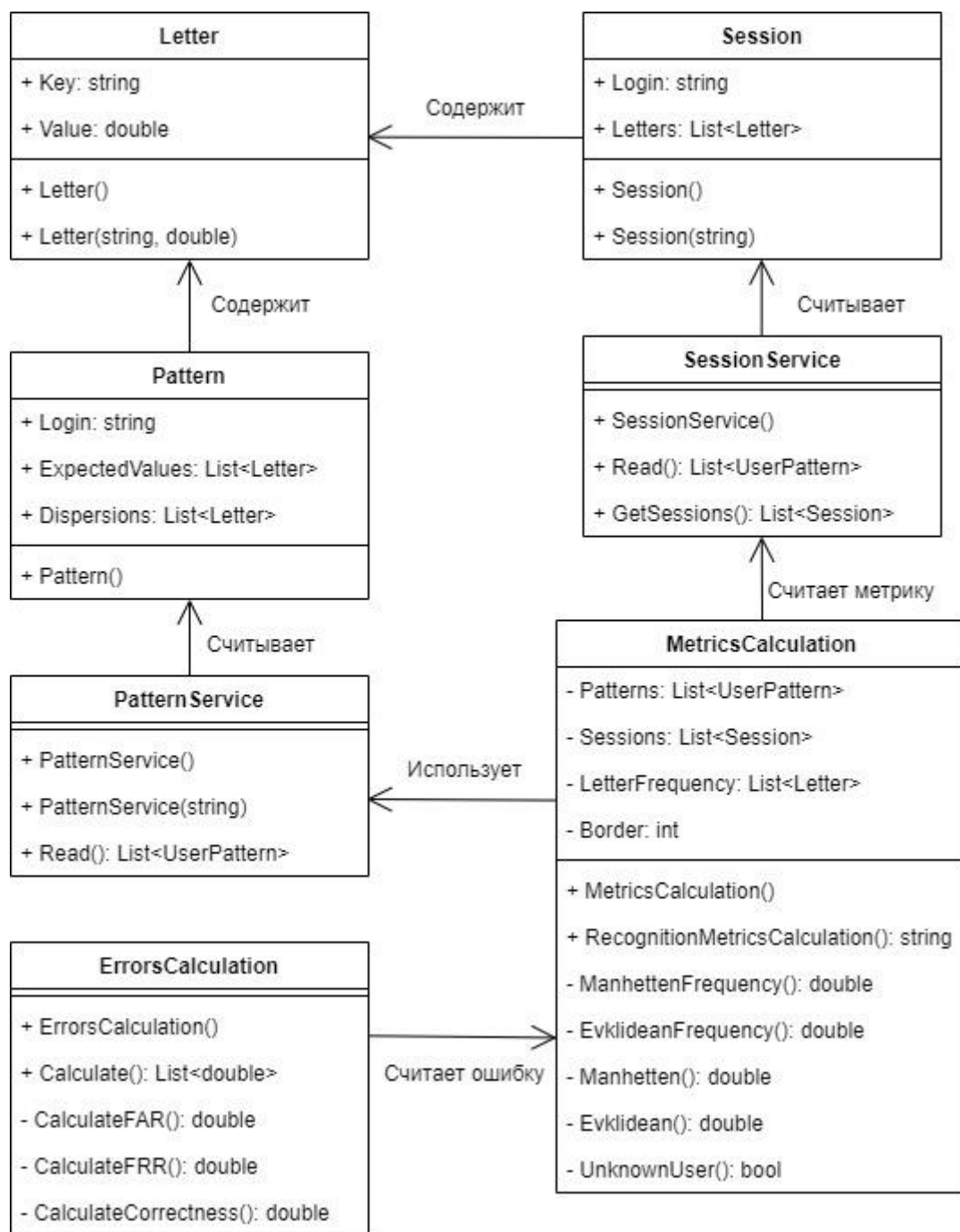


Рисунок 3.9 – Диаграмма классов

Класс `PatternService` предназначен для получения шаблонов сессий из хранилища данных. Шаблоны сессий записываются в объекты класса `Pattern`.

Класс `SessionService` предназначен для получения сессий из хранилища данных. Шаблоны сессий записываются в объекты класса `Session`.

Класс `Letter` используется для описания буквы, нажатой пользователем, и временем удержания клавиши.

Класс MetricsCalculation используется для решения задачи идентификации путем расчета различных методов.

Класс ErrorsCalculation предназначен для расчета ошибок методов.

На рисунке 3.10 изображена диаграмма процесса «Идентификация».

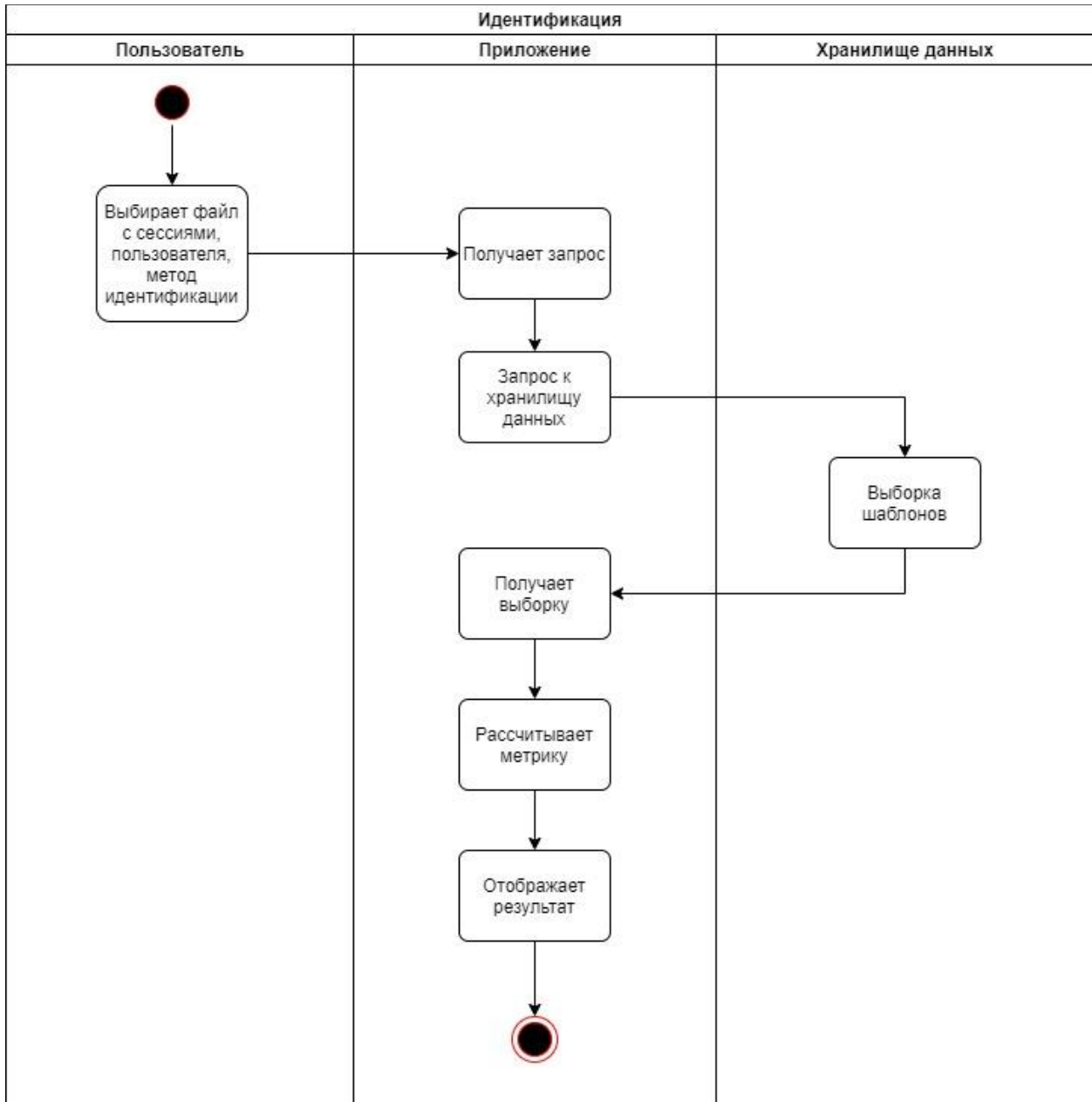


Рисунок 3.10 – Диаграмма процесса «Идентификация»

3.3.2 Описание интерфейса

В состав приложения входят три окна (рисунок 3.11).

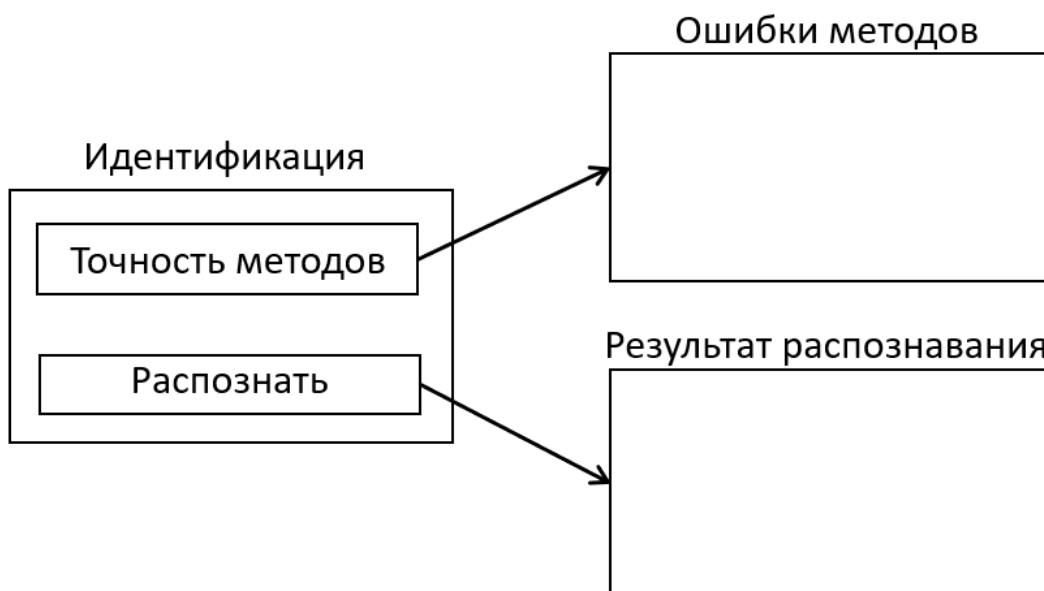


Рисунок 3.11 – Структура окон приложения

На рисунке 3.12 отображено основное окно приложения.

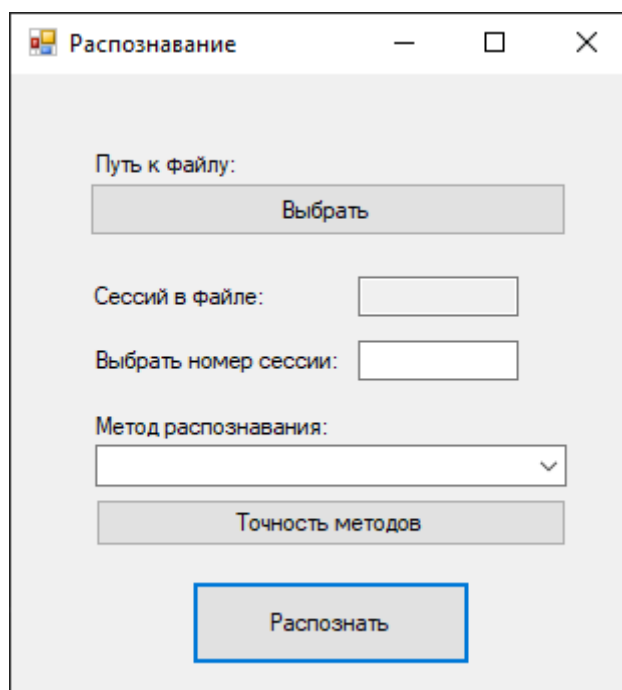


Рисунок 3.12 – Основное окно приложения

Окно приложения имеет в своем составе:

1. Кнопка «Выбрать», предназначенная для выбора файла с сессиями.
2. Поле «Сессий в файле». Отображает количество сессий в выбранном файле.
3. Поле «Номер сессии». Предназначено для ввода номера сессии для распознавания.

4. Поле «Метод распознавания». Поле предназначено для выбора метода распознавания.

5. Кнопка «Точность методов». При нажатии пользователю отображается форма «Ошибки методов».

6. Кнопка «Распознать». Кнопка предназначена для запуска режима распознавания сессии пользователя.

Для запуска процесса распознавания пользователю необходимо выбрать файл с сессиями, затем выбрать номер сессии, выбрать метод распознавания и нажать на кнопку «Распознать» (рисунок 3.13).

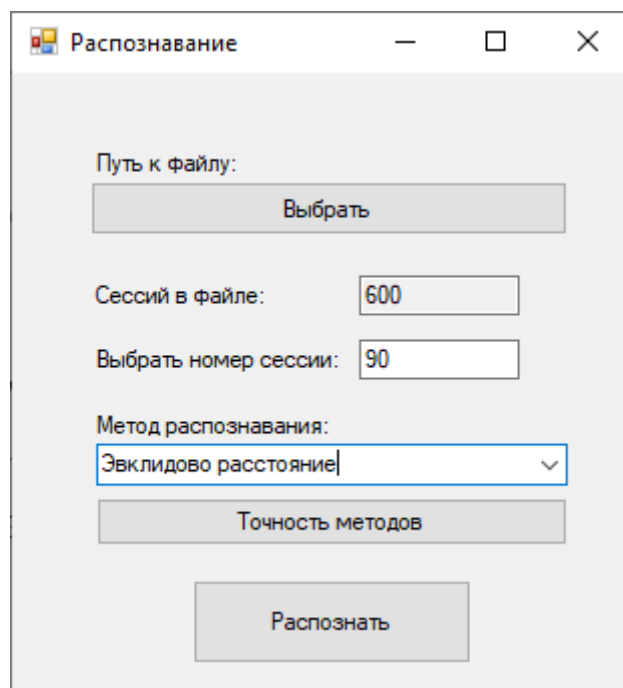


Рисунок 3.13 – Ввод значений для запуска процесса распознавания

Результат распознавания сессии изображен на рисунке 3.14

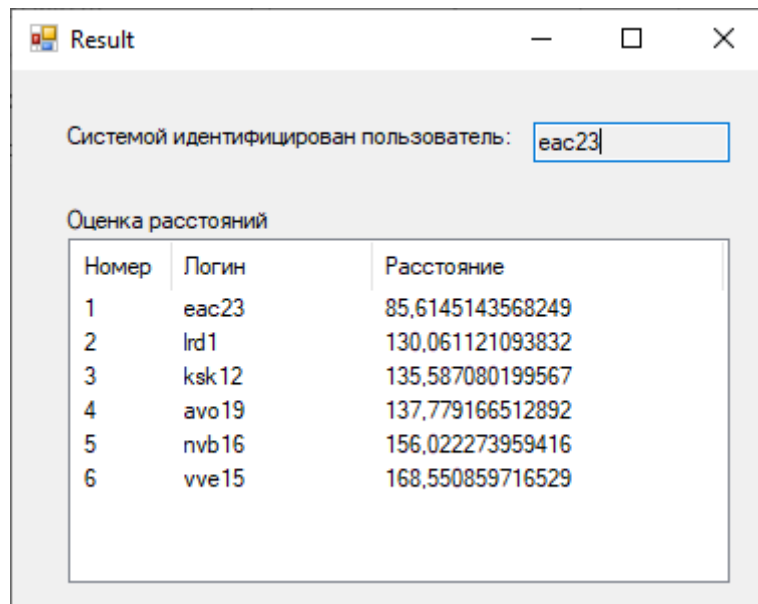


Рисунок 3.14 –Результат распознавания сессии

Для просмотра точности методов распознавания пользователю необходимо нажать на кнопку «Точность методов». После нажатия появится окно (рисунок 3.15).

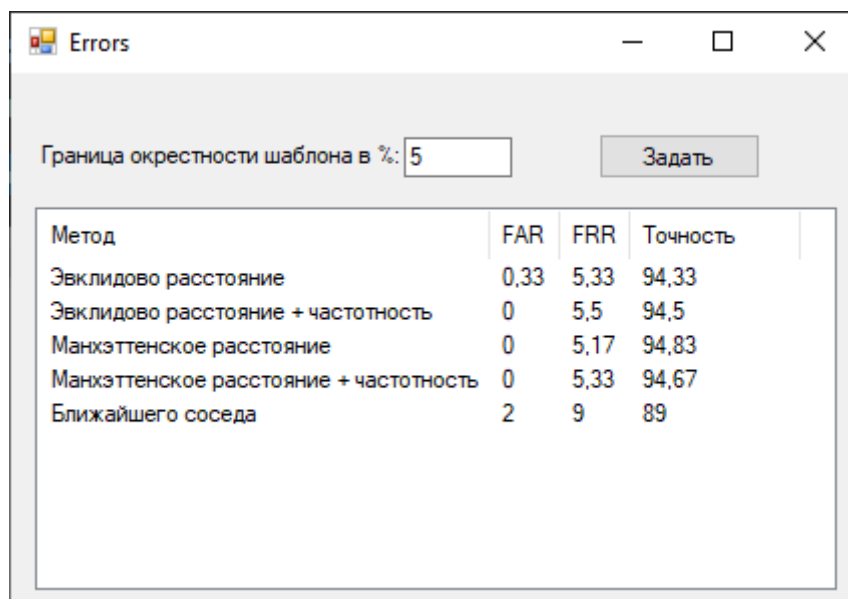


Рисунок 3.15 – Окно «Точность методов» при значении окрестности 5%

На рисунке 3.16 изображено, как задать «Границу окрестности». Для этого необходимо в поле «Граница окрестности шаблона» ввести число от 0 до 100 и нажать на кнопку «Задать».

Метод	FAR	FRR	Точность
Эвклидово расстояние	0,67	0	99,33
Эвклидово расстояние + частотность	0,17	0,67	99,17
Манхэттенское расстояние	0	0	100
Манхэттенское расстояние + частотность	0,17	0,33	99,5
Ближайшего соседа	3,5	3	93,5

Рисунок 3.16 – Окно «Точность методов» при значении границы окрестности 8%

3.3.3 Анализ результатов

В ходе исследования были проанализированы методы распознавания пользователей. Для анализа задавались различные значения границ окрестности шаблона.

В таблицах 2 и 3 приведены ошибки методов распознавания, при различных значениях границы окрестности.

Таблица 2 – Сравнительный анализ показателей эффективности идентификации при значении границы окрестности 5%.

№	Метод	FAR	FRR	Точность
1	Эвклидово расстояние	0.33	5.33	94.33
2	Эвклидово расстояние + частотность букв русского языка	0	5.5	94.5
3	Манхэттенское расстояние	0	5.17	94.83
4	Манхэттенское расстояние + частотность букв русского языка	0	5.33	94.67
5	Ближайшего соседа	2	9	89

Таблица 3 – Сравнительный анализ показателей эффективности идентификации при значении границы окрестности 8%.

№	Метод	FAR	FRR	Точность
1	Эвклидово расстояние	0.67	0	99.33
2	Эвклидово расстояние + частотность букв русского языка	0.17	0.67	99.17
3	Манхэттенское расстояние	0	0	100
4	Манхэттенское расстояние + частотность букв русского языка	0.17	0.33	99.5
5	Ближайшего соседа	3.5	3	93.5

В ходе исследования выяснилось, что частотность букв русского алфавита незначительно повышает точность методов распознавания.

Результаты исследования зависимости чувствительности алгоритмов, ошибок и точности изображены в таблице 4.

Таблица 4 – Результаты исследования

Чувствительность	FAR	FRR	Точность
↑	↑	↓	↑
↓	↓	↑	↓

Значимость ошибок FAR и FRR зависит от конкретной прикладной задачи. Это не означает, что точность идентификации является менее важным критерием для системы идентификации. Однако, в данном случае риск пропуска «шпиона» намного выше, чем риск отказа в доступе к системе зарегистрированному пользователю. Проанализировав вышеописанные факты, было принято решение установить порог чувствительности системы 10%, как самый оптимальный.

4 Финансовый менеджмент

4.1 Предпроектный анализ

4.1.1 Потенциальные потребители результатов исследования

Система распознавания пользователя по клавиатурному почерку предназначена для защиты информации. Поэтому целевой рынком для неё являются компании, заинтересованные в контроле доступа к информации. Таким компаниям включают: военные предприятия, производственные предприятия, софтверные компании, торговые компании, научно-исследовательские центры. Сегментацию рынка возможно провести по следующим критериям.

Размер предприятия

1. Крупные;
2. Средние;
3. Малые;

Отраслевая принадлежность предприятия

1. Промышленность;
2. Финансовые услуги;
3. Торговля;
4. Транспорт;
5. Сфера услуг;

Степень секретности информации

1. Высокая;
2. Средняя;
3. Низкая;

Используемая операционная система

1. Windows OS;
2. Mac OS;
3. Linux OS;

В качестве наиболее значимых были выбраны следующие критерии: степень секретности информации и размер предприятия. В результате анализа конкурентных программных продуктов было установлено, что относительно

небольшая конкуренция наблюдается в случае, если предприятие малое и степень секретности информации низкая либо средняя.

4.1.2 Анализ конкурентных технических решений

Для оценки конкурентоспособности разработки проводится анализ существующих решений, позволяющих анализировать и распознавать клавиатурный почерк. Для сравнительного анализа были выбраны:

1. Облачное приложение «CleverControl» для контроля активности сотрудников на компьютерах в режиме реального времени

2. Система мониторинга работы подчиненных «StaffCop», включающая в себя модуль для анализа клавиатурного почерка

Сравнение технических и экономических характеристик этих продуктов представлено в таблице 1. «CleverControl» обозначен К1, а «StaffCop» - К2.

Таблица 5 – Оценочная карта сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентно-способность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1.Удобство в эксплуатации	0,18	4	1	5	0,72	0,18	0,9
2.Возможность подключения в сеть ЭВМ	0,11	4	1	5	0,44	0,11	0,55
3.Потребность в ресурсах	0,07	5	5	2	0,35	0,35	0,14
4.Функциональные возможности	0,15	3	1	5	0,45	0,15	0,75
5.Быстродействие	0,1	5	5	4	0,5	0,5	0,4
Экономические критерии оценки ресурсоэффективности							
1.Возможность доработки	0,06	5	3	4	0,3	0,18	0,24
2.Цена	0,2	4	5	1	0,8	1	0,2
3.Уровень проникновения на рынок	0,1	1	2	4	0,1	0,2	0,4
4.Обслуживание после продажи	0,02	3	1	5	0,06	0,02	0,1

5.Предполагаемый срок эксплуатации	0,01	5	3	4	0,05	0,03	0,04
Итого	1				3,77	2,72	3,72

Таким образом, разрабатываемая система распознавания клавиатурного почерка имеет преимущества перед аналогами по следующим критериям:

1. Цена;
2. Потребность в ресурсах памяти;
3. Быстродействие;
4. Предполагаемый срок эксплуатации;
5. Возможность подключения в сеть ЭКМ;
6. Удобство в эксплуатации.

Недостатками системы являются:

1. Уровень проникновения на рынок;
2. Функциональные возможности;
3. Послепродажное обслуживание.

В результате анализа было установлено, что система является более конкурентноспособной, чем К1 и К2. Следовательно, более целесообразно проводить дальнейшую разработку.

4.1.3 SWOT-анализ

SWOT-анализ используется для выявления сильных и слабых сторон проекта, а также его возможностей и угроз выполнения.

На первом этапе SWOT-анализа были описаны сильные и слабые стороны проекта, возможности и угрозы реализации. Матрица SWOT-анализа представлена в таблице 6.

Таблица 6 – SWOT-анализ

	<p>Сильные стороны научно-исследовательского проекта:</p> <p>С1. Простота эксплуатации</p> <p>С2. Низкая стоимость разработки</p>	<p>Слабые стороны научно-исследовательского проекта:</p> <p>Сл1. Невысокая точность алгоритма распознавания клавиатурного почерка</p>
--	--	--

	С3. Централизованное хранение данных С4. Низкие требования к аппаратно-программному обеспечению С5. Удобный интерфейс С6. Графическое представление данных	Сл2. Отсутствие кроссплатформенности Сл3. Длительная разработка
Возможности: В1. Реализация новых функций В2. Повышение отказоустойчивости программы В3. Увеличение спроса на продукт В4. Расширение команды разработчиков для ускорения реализации и поддержки продукта В5. Реализация версий программы для Linux и MacOS		
Угрозы: У1. Увеличение конкуренции У2. Прекращение поддержки руководителей проекта У3. Отсутствие интереса к продукту на рынке		

Второй этап SWOT-анализа включает выявление соответствий между сильными и слабыми сторонами проекта и окружающей средой. Интерактивные матрицы соответствия представлены в таблицах 7-10.

Таблица 7 – Интерактивная матрица соответствия сильных сторон и возможностей

		Сильные стороны проекта					
		Сл1	Сл2	Сл3	Сл4	С5	С6
Возможности проекта	В1	-	+	+	0	-	+
	В2	-	+	+	+	-	-
	В3	+	-	+	+	+	+
	В4	-	+	-	-	-	-

	B5	0	+	-	+	-	-
--	----	---	---	---	---	---	---

Таблица 8 – Интерактивная матрица соответствия сильных сторон и угроз

		Сильные стороны проекта					
Угрозы проекта		Сл1	Сл2	Сл3	Сл4	С5	С6
	У1	-	-	-	-	0	-
	У2	-	-	-	-	-	-
	У3	-	+	-	-	-	-

Таблица 9 – Интерактивная матрица соответствия слабых сторон и возможностей

		Слабые стороны проекта		
Возможности проекта		Сл1	Сл2	Сл3
	B1	+	-	+
	B2	-	-	+
	B3	-	-	-
	B4	+	-	+
	B5	-	+	+

Таблица 10 – Интерактивная матрица соответствия слабых сторон и угроз

		Слабые стороны проекта		
Угрозы проекта		Сл1	Сл2	Сл3
	У1	+	+	+
	У2	+	-	+
	У3	+	+	-

Третий этап включает в себя составление итоговой матрицы SWOT-анализа на основе полученной таблицы SWOT-анализа и интерактивных таблиц (таблица 11).

Таблица 11 – Итоговая матрица SWOT-анализа

	<p>Сильные стороны научно-исследовательского проекта:</p> <p>С1. Простота эксплуатации</p> <p>С2. Низкая стоимость разработки</p> <p>С3. Централизованное хранение данных</p> <p>С4. Низкие требования к аппаратно-программному обеспечению</p> <p>С5. Удобный интерфейс</p>	<p>Слабые стороны научно-исследовательского проекта:</p> <p>Сл1. Невысокая точность алгоритма распознавания клавиатурного почерка</p> <p>Сл2. Отсутствие кроссплатформенности</p> <p>Сл3. Длительная разработка</p>
--	---	--

	С6. Графическое представление данных	
Возможности: В1. Реализация новых функций В2. Повышение отказоустойчивости программы В3. Увеличение спроса на продукт В4. Расширение команды разработчиков для ускорения реализации и поддержки продукта В5. Реализация версий программы для Linux и MacOS	1. В1С2С3С6 – Простота расширения функционала системы. 2. В2С2С3С4 – Простота и низкая стоимость изменения каналов связи. 3. В3С1С3С4С5С6 – Широкие возможности для увеличения спроса. 4. В4С2 – Ускорение разработки. 5. В5С2С4 – Расширение рынка сбыта.	1. В1Сл1Сл3 – Необходимость доработки и оптимизации алгоритма. 2. В2Сл3 – Модификация приложения требует времени. 3. В4Сл1Сл3 – Новые разработчики должны сначала исследовать существующий код и алгоритм. 4. В5Сл2Сл3 – Реализация поддержки новой ОС потребует
Угрозы: У1. Увеличение конкуренции У2. Прекращение поддержки руководителей проекта У3. Отсутствие интереса к продукту на рынке	1. У3С2 – Недостаточно системный подход к разработке ПО	1. У1Сл1Сл2Сл3 – Конкуренты смогут разработать ПО сходного функционала более быстро и качественно. 2. У2Сл1Сл3 – Руководитель проекта недоволен реализацией. 3. У3Сл1Сл2 – Продукт не удовлетворяет ожидания пользователей.

4.2 Планирование научно-исследовательских работ

4.2.1 Структура работ в рамках научного исследования

Для организации и систематизации работы выпускника был сформирован план работ. Данный этап обеспечил своевременное и эффективное выполнение задания выпускной квалификационной работы.

Для осуществления разработки, был сформирован ряд работ и назначены должности исполнителей для каждого этапа работы (таблица 12).

Таблица 12 – Перечень работ по проекту

№ работы	Наименование работы	Исполнители работы
----------	---------------------	--------------------

1	Выбор научного руководителя бакалаврской работы	Затеев Р.П..
2	Составление и утверждение темы бакалаврской работы	Кочегурова Е.А., Затеев Р.П..
3	Составление календарного плана-графика выполнения бакалаврской работы	Кочегурова Е.А.
4	Подбор и изучение литературы по теме бакалаврской работы	Затеев Р.П..
5	Анализ предметной области	Затеев Р.П..
6	Разработка методики и алгоритма распознавания	Затеев Р.П..
7	Проектирование системы распознавания пользователя по клавиатурному почерку	Затеев Р.П..
8	Сбор и моделирование тестовых данных	Затеев Р.П..
9	Разработка системы распознавания пользователя по клавиатурному почерку	Затеев Р.П..
10	Тестирование системы	Затеев Р.П..
11	Оценка эффективности полученных результатов	Затеев Р.П..
12	Согласование выполненной работы с научным руководителем	Кочегурова Е.А., Затеев Р.П..
13	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Затеев Р.П..
14	Подведение итогов, оформление работы	Затеев Р.П..

4.2.2 Определение трудоемкости выполнения работ

Для того чтобы определить трудоемкость работ, используются следующие показатели:

- Ожидаемое значение трудоемкости,
- Продолжительность каждой работы,
- Продолжительность выполнения i -ой работы в календарных днях,
- Коэффициент календарности.

Расчет ожидаемого значения продолжительности работ $t_{ож}$ осуществляется согласно формуле:

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5}, \quad (4.1)$$

где t_{min} – минимально возможная трудоемкость i -ой работы, чел.-дни;

t_{max} – максимально возможная трудоемкость i -ой работы, чел.-дни.

Далее определяется продолжительность каждой работы в рабочих днях T_{pi} , которая учитывает параллельность выполнения работ несколькими исполнителями:

$$T_{pi} = \frac{t_{ожi}}{Ч_i}, \quad (4.2)$$

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

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

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

4.2.3 Разработка графика проведения научного исследования

Для построения графика, осуществляется перевод длительности каждого из этапов работ из рабочих дней в календарные дни по следующей формуле:

$$T_{ki} = T_{pi} \cdot k_{кал}, \quad (4.3)$$

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

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

$k_{кал}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}, \quad (4.4)$$

где $T_{кал}$ – количество календарных дней в году,

$T_{вых}$ – количество выходных дней в году,

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

Согласно производственному календарю (для 6-дневной рабочей недели) в 2020 году 366 календарных дней, из них 97 выходных или праздничных дней, следовательно, $k_{кал} = 1,36$. Расчеты по трудоемкости выполнения работ представлены в таблице 13. Диаграмма Ганта, построенная по рассчитанным показателям, представлена на рисунке 4.1.

Таблица 13 – Временные показатели осуществления разработки

№	Наименование работы	Исполнители работы	Трудоемкость работ, чел.-дни			Длительность работ, дни	
			t_{min}	t_{max}	$t_{ож}$	T_p	T_k
1	Выбор научного руководителя бакалаврской работы	Затеев Р.П.	1	2	1,4	1	1
2	Составление и утверждение темы бакалаврской работы	Затеев Р.П.	3	4	3,4	2	3
		Кочегурова Е.А.	3	4	3,4	2	3
3	Составление календарного плана-графика выполнения бакалаврской работы	Кочегурова Е.А.	2	4	2,8	3	4
4	Подбор и изучение литературы по теме бакалаврской работы	Затеев Р.П.	5	8	6,2	6	8
5	Анализ предметной области	Затеев Р.П.	4	7	5,2	5	7
6	Разработка методики и алгоритма распознавания	Затеев Р.П.	10	12	10,8	11	15
7	Проектирование системы распознавания пользователя по клавиатурному почерку	Затеев Р.П..	10	12	10,8	11	15
8	Сбор и моделирование тестовых данных	Затеев Р.П.	10	12	10,8	11	15
9	Разработка системы распознавания пользователя по клавиатурному почерку	Затеев Р.П.	16	18	16,8	17	23
10	Тестирование системы	Затеев Р.П.	10	14	11,6	12	16
11	Оценка эффективности полученных результатов	Затеев Р.П.	5	7	5,8	6	8
12	Согласование выполненной работы с научным руководителем	Затеев Р.П.	3	5	3,8	2	3
		Кочегурова Е.А.	3	5	3,8	2	3
13	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Затеев Р.П.	5	7	5,8	6	8
14	Подведение итогов, оформление работы	Затеев Р.П.	3	4	3,4	3	4
	Итого	Кочегурова Е.А.			7	10	
		Затеев Р.П.			93	126	

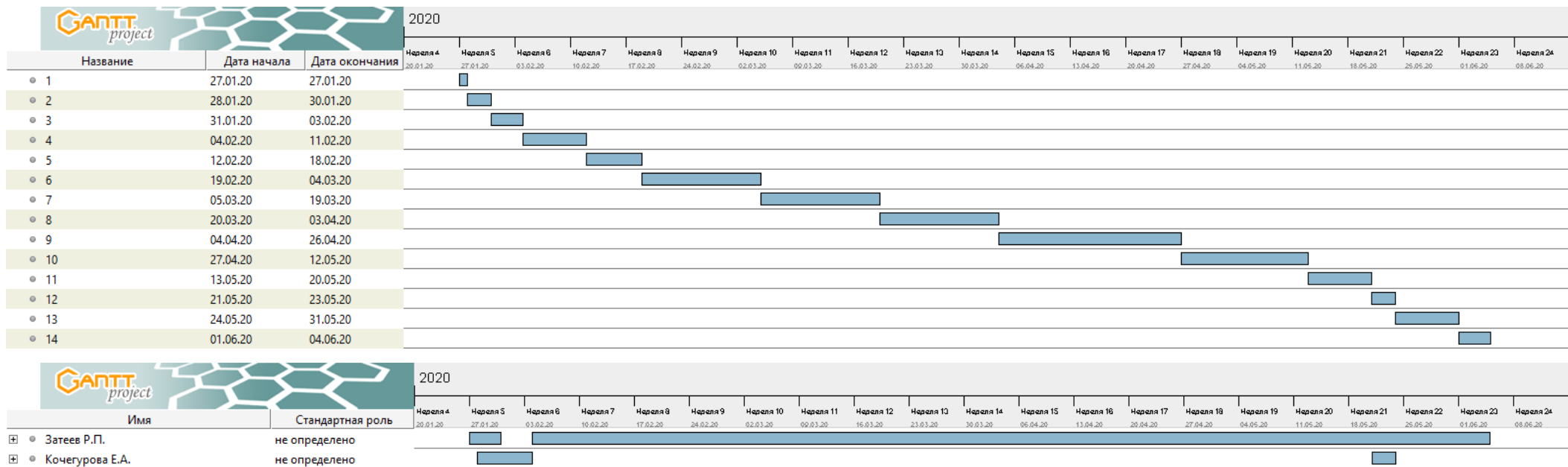


Рисунок 4.1 – Диаграмма Гантта: этапы выполнения работ(сверху), загруженность исполнителей(снизу).

4.2.4 Бюджет научно-технического исследования

4.2.4.1 Расчет материальных затрат

Данная статья затрат включает в себя затраты на приобретение сырья, материалов, полуфабрикатов и комплектующих со стороны. Также в эту статью включаются транспортные расходы, равные 15 % от общей стоимости материальных затрат.

Общая сумма материальных затрат включает в себя только затраты на канцелярские принадлежности (300 руб.), для которых не учитываются транспортные расходы.

Таким образом, общая сумма материальных затрат составляет 300 руб.

4.2.4.2 Расчет затрат на специальное оборудование

В данную статью затрат входят суммы, необходимые на обеспечение амортизации используемого оборудования.

Расчет амортизации персонального компьютера, используемого при написании работы: первоначальная стоимость персонального компьютера составляет 54000 рублей; срок полезного использования для офисных машин – 3 года; планируется использовать персональный компьютер для написания ВКР в течение 4 месяцев.

Тогда:

- норма амортизации:

$$A_n = \frac{1}{n} \cdot 100\% = \frac{1}{3} \cdot 100\% = 33,33 \%, \quad (4.5)$$

- годовые амортизационные отчисления:

$$A_g = 54000 \cdot 0,33 = 17820 \text{ руб.}, \quad (4.6)$$

- ежемесячные амортизационные отчисления:

$$A_m = \frac{17820}{12} = 1485 \text{ руб.}, \quad (4.7)$$

- итоговая сумма амортизации основных средств:

$$A = 1485 \cdot 4 = 5940 \text{ руб.} \quad (4.8)$$

Итоговая сумма затрат на амортизацию составила 5940 рублей.

4.2.4.3 Основная заработная плата исполнителей темы

Месячный оклад руководителя ТПУ с должностью доцента и степенью кандидата наук составляет 33664 рубля (без учета районного коэффициента, но с учетом премиальных и надбавок), для студента был взят оклад ассистента без научной степени – 21760 рублей.

В таблице 14 показаны количества календарных, нерабочих и праздничных дней, дней, пришедшихся на потерю рабочего времени, и действительный годовой фонд рабочего времени F_d .

Таблица 14 – Баланс рабочего времени (для 6-дневной недели)

Показатели рабочего времени	Дни
Календарные дни	366
Нерабочие дни (праздники/выходные)	97
Потери рабочего времени (отпуск/невыходы по болезни)	56
Действительный годовой фонд рабочего времени	213

Количество месяцев работы без отпуска принимается за $M = 10,4$ (с учетом длительности отпуска в 48 дней). Тогда, зная месячную заработную плату, можно рассчитать среднедневную заработную плату:

$$Z_{\text{рук}}^{\text{рук}} = \frac{Z_M \cdot M}{F_d} = \frac{33664 \cdot 10,4}{213} = 1643,69 \text{ руб.}, \quad (4.9)$$

$$Z_{\text{студ}}^{\text{студ}} = \frac{Z_M \cdot M}{F_d} = \frac{21760 \cdot 10,4}{213} = 1062,46 \text{ руб.} \quad (4.10)$$

Расчет основной заработной платы осуществляется по формуле:

$$Z_{\text{осн}} = Z_{\text{дн}} \cdot T_p \cdot (1 + K_{\text{пр}} + K_d) \cdot K_p, \quad (4.11)$$

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

T_p – продолжительность работ, выполняемых работником, раб. дни,

$K_{\text{пр}}$ – премиальный коэффициент,

K_d – коэффициент доплат и надбавок,

K_p – районный коэффициент.

Результаты соответствующих расчетов приведены в таблице 15.

Таблица 15 – Расчет основной заработной платы

Исполнители	$Z_{доп}$, руб.	$K_{пр}$	$K_{д}$	$K_{р}$	T_p	$Z_{осн}$, руб.
Студент	1062,46	0,3	0,2	1,3	93	192677,12
Научный руководитель	1643,69	0,3	0,2	1,3	7	22436,37
Итого						215113,49

4.2.4.4 Дополнительная заработная плата исполнителей темы

Зная основную заработную плату, можно рассчитать дополнительную заработную плату в размере 12 % от основной:

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

где $k_{доп}$ – коэффициент дополнительной заработной платы,

$Z_{осн}$ – основная заработная плата.

Таблица 16 – Расчет дополнительной заработной платы

Исполнители	$k_{доп}$	$Z_{осн}$, руб.	$Z_{доп}$, руб.
Студент	0,12	192677,12	23121,25
Научный руководитель	0,12	22436,37	2692,36
Итого			25813,61

4.2.4.5 Отчисления во внебюджетные фонды (страховые отчисления)

Отчисления во внебюджетные фонды рассчитываются как:

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

где $k_{внеб}$ – коэффициент внебюджетных фондов, для образовательных учреждений используется пониженная ставка, принятая в 2019 году в размере 30 %;

$Z_{осн}$ – основная заработная плата;

$Z_{доп}$ – дополнительная заработная плата.

Таблица 17 – Расчет страховых отчислений

Исполнители	$k_{внеб}$	$Z_{доп}$	$Z_{осн}$	$Z_{внеб}$
Студент	0,30	23121,25	192677,12	64739,51
Научный руководитель	0,30	2692,36	22436,37	9423,27
Итого				74162,78

4.2.4.6 Накладные расходы

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

$$Z_{\text{накл}} = k_{\text{накл}} \cdot (\text{сумма статей расходов}), \quad (4.14)$$

где $k_{\text{накл}}$ – коэффициент накладных расходов, принятый за 16 %.

Таблица 18 – Расчет накладных расходов

Статьи затрат	Сумма, руб.
Материальные затраты	300
Затраты на амортизацию	5940
Затраты на основную заработную плату	215113,49
Затраты на дополнительную заработную плату	25813,61
Затраты на отчисления во внебюджетные фонды	74162,78
Накладные расходы	51412,78

4.2.4.7 Формирование бюджета затрат проекта разработки

Рассчитанные величины затрат научно-исследовательской работы являются основой для формирования бюджета затрат проекта. Результаты составления итогового бюджета разработки представлены в таблице 19.

Таблица 19 – Бюджет затрат на разработку

Наименование	Сумма, руб.	Удельный вес, %
Материальные затраты	300	0,08
Затраты на амортизацию	5940	1,59
Затраты на основную заработную плату	215113,49	57,71
Затраты на дополнительную заработную плату	25813,61	6,92
Затраты на отчисления во внебюджетные фонды	74162,78	19,9
Накладные расходы	51412,78	13,79
Общий бюджет	372742,63	100

4.3 Вывод по разделу

В ходе осуществления оценки конкурентоспособности разработки было установлено, что разработанная система обладает высоким уровнем конкурентоспособности; общая длительность разработки составляет 129 календарных дней (период с 27.01.2020 по 04.06.2020); общий бюджет разработки был оценен в 372742,63 рубля.

В дальнейшем, разработанная система может быть использована учебными заведениями, либо коммерческими организациями для обеспечения информационной безопасности и улучшения эффективности своей работы.

5 Социальная ответственность

Объектом исследования данной ВКР является система распознавания пользователя по клавиатурному почерку. Как разработка системы, так и её эксплуатация происходит в офисном помещении. Офис, а также находящаяся в нем компьютерная и оргтехника, оказывают неблагоприятное воздействие на окружающую среду. Кроме того, при несоблюдении санитарных норм и правил, возможно, негативное влияние на здоровье сотрудников офиса.

Данный раздел посвящен вопросам производственной безопасности и гигиене труда, соблюдению санитарных норм и защиты сотрудников от негативного воздействия среды. Рассматриваются меры по охране окружающей среды и ресурсосбережению. Предлагается ряд решений для исключения возникновения несчастных случаев при разработке и эксплуатации системы распознавания клавиатурного почерка.

Также исследуются правовые вопросы обеспечения безопасности, связанные с использованием разработанной системы.

5.1 Правовые и организационные вопросы обеспечения безопасности

5.1.1 Правовые и организационные мероприятия по обеспечению безопасности

Функции государственного надзора и контроля обеспечения безопасности осуществляются специально уполномоченными государственными органами и инспекциями согласно федеральным законам.

Для защиты прав сотрудников на труд в условиях, соответствующих принятым правилам и нормам, на территории Российской Федерации действуют следующие организации.

- Федеральная инспекция труда;
- Государственная экспертиза условий труда Федеральная служба по труду и занятости населения;
- Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека.

ТК РФ и СанПиН 2.2.2/2.4.1340-03 [12] устанавливаются следующие правила и вводятся рекомендации, касающиеся работы сотрудников всех организаций на территории РФ:

1. Нормальная продолжительность рабочего времени не должна превышать 40 часов в неделю.

2. Продолжительность непрерывной работы за компьютером без перерыва не должна превышать 1 час;

3. Рекомендуется делать перерывы в работе за компьютером продолжительностью 10-15 минут через каждые 45-60 минут работы. Во время перерывов целесообразно выполнять комплексы упражнений, в том числе упражнения для глаз, а также проветривать помещения;

4. Не рекомендуется работать за компьютером более 6 часов за смену.

Для того чтобы рабочие места сотрудников соответствовали нормам, должен осуществляться производственный контроль и надзор внутри предприятия. Кроме того, предприятия следят за характеристиками используемой аппаратуры, персональных компьютеров и комплектующих.

5.1.2 Специальные правовые нормы трудового законодательства

Государственный надзор и контроль в организациях независимо от организационно-правовых форм и форм собственности осуществляют специально уполномоченные на то государственные органы и инспекции в соответствии с федеральными законами. Согласно трудовому кодексу РФ:

- продолжительность рабочего дня не должна превышать 40 часов в неделю;
- во время регламентированных перерывов целесообразно выполнять комплексы упражнений и осуществлять проветривание помещения.

Существуют также специализированные органы, осуществляющие государственный контроль и надзор в организациях на предмет соблюдения существующих правил и норм.

К таким органам относятся:

- Федеральная инспекция труда;

- Государственная экспертиза условий труда Федеральной службы по труду и занятости населения;
- Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека и др.

5.1.3 Организационные мероприятия при компоновке рабочей зоны

Рабочее место для выполнения работ сидя организуют при легкой работе, не требующей свободного передвижения работающего, а также при работе средней тяжести в случаях, обусловленных особенностями технологического процесса.

Конструкция рабочего места и взаимное расположение всех его элементов должны соответствовать антропометрическим, физиологическим и психологическим требованиям, а также характеру работы.

5.2 Производственная безопасность

Вредные и опасные факторы, воздействующие на сотрудника, устанавливаются согласно ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация».

Опасные и вредные факторы при выполнении работ по разработке и эксплуатации программы:

Таблица 20 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Разрабо тка	Эксплу атация	
1. Недостаточная освещенность рабочей зоны	+		Освещение, шум, статическое электричество, психофизиологические факторы: СанПиН 2.2.2/2.4.1340-03 ТОИ П-45-084-01 Микроклимат: СанПиН 2.2.4.548-96 СанПиН 2.2.2/2.4.1340-03 Электрический ток, короткое замыкание: «Правила устройства электроустановок» утвержденные Главтехуправлением,
2. Повышенный уровень шума на рабочем месте	+	+	
3. Отклонение параметров микроклимата на рабочем месте.	+	+	
4. Нервно-психические перегрузки.	+	+	

			Госэнергонадзором Минэнерго СССР 05.10.1979 г. ГОСТ 12.1.033-81 «Правила технической эксплуатации электроустановок потребителей», утвержденные приказом Минэнерго России от 13.01.2003 г. №6 «Межотраслевые правила охраны труда (правила безопасности) при эксплуатации электроустановок» (ПОТ РМ 016-2001), утвержденные Постановлением Минтруда России от 05.01.2001 г №3
5.Перенапряжение зрительных анализаторов.	+	+	
6.Повышенный уровень электромагнитных излучений.	+	+	
7 Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека	+	+	
8	+	+	
9.Повышенный уровень статического электричества.	+	+	

5.2.1 Недостаточная освещенность рабочей зоны

На время проектирования, а также эксплуатации приложения за персональным компьютером к негативным факторам можно отнести повышенные уровни излучения, а также увеличенную нагрузку на зрительные органы.

Требования к освещению установлены в СНиП 23-05-95.

Таблица 21 – Требования к освещению на рабочих местах, оборудованных ПЭВМ

Характеристика зрительной работы		очень высокой точности	
Наименьший или эквивалентный размер объекта различия, мм		(0,15 - 0,3)	
Разряд зрительной работы		A	
Подразряд зрительной работы		2	
Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %		менее 70	
Естественное освещение	освещенность на рабочей поверхности от системы общего освещения, лк	(300 – 500)	
	цилиндрическая освещенность, лк	100	
	показатель дискомфорта M	не более 15	
	коэффициент пульсации освещенности K_p , %	10	
Искусственное освещение	КЕО e_n , %, при	верхнем или комбинированном	3,5
		боковом	1,2

5.2.2 Превышение уровня шума

Повышенный уровень шума характерен не только для промышленных предприятий, но и для бытовых приборов. Шумовой фон в современных городах сопоставим с допустимым максимумом.

Интенсивность шума – главная характеристика. Она тождественна громкости и измеряется в децибелах (дБ).

В производственных помещениях при выполнении основных или вспомогательных работ с использованием ПЭВМ уровни шума на рабочих местах не должны превышать предельно допустимых значений, установленных для данных видов работ в соответствии с действующими санитарно-эпидемиологическими нормативами. СанПиН 2.2.2/2.4.1340-03 [12] установлены допустимые значения уровней звукового давления, создаваемого ПЭВМ [16].

Таблица 22 – Допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ

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

Повышенный шум на рабочем месте оказывает вредное влияние на организм и приводит к неблагоприятным изменениям в его органах и системах. Длительное воздействие шума способно привести к потере слуха, болезням сердечно-сосудистой системы, артериальной гипертензии, нарушениям в работе нервной системы и т.д.

5.2.3 Отклонение параметров микроклимата

К числу показателей микроклимата, которые измеряют для его оценки, относятся следующие:

- скорость движения воздуха;
- относительная влажность воздуха;
- температура;

- интенсивность теплового облучения.

В производственных помещениях, в которых работа с использованием ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты управления, залы вычислительной техники и др.) и связана с нервно-эмоциональным напряжением, должны обеспечиваться оптимальные параметры микроклимата для категории работ 1а и 1б в соответствии с действующими санитарно-эпидемиологическими нормативами микроклимата производственных помещений.

Содержание вредных химических веществ в производственных помещениях, в которых работа с использованием ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты управления, залы вычислительной техники и др.), не должно превышать предельно допустимых концентраций загрязняющих веществ в атмосферном воздухе населенных мест в соответствии с действующими гигиеническими нормативами.

Также указывается, что в помещениях с ПЭВМ должна ежедневно проводиться влажная уборка.

Уставленные гигиенические нормативы для помещений с ВДТ и ПЭВМ для категории работы 1б приведены в таблице.

Таблица 23 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений категории 1б (СанПиН 2.2.4.548-96 [14])

Период года	Категория работ по уровню энергозатрат	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	1б (140-174 Вт)	(21 – 23)	(20 – 24)	(40 – 60)	0,1
Теплый		(22 – 24)	(21 – 25)	(40 – 60)	0,1

5.2.4 Нервно-психические нагрузки

Среди причин, ведущих к тяжелым заболеваниям, нужно указать на нервные перегрузки, отрицательные психологические раздражители.

Перегрузки возникают при неравномерном распределении времени при работе и отдыхе. Это приводит к утомлению, что является реакцией организма

на физические и нервно-психические перегрузки. Вследствие накопления усталости постепенно развивается переутомление, что относится к состоянию предболезни. Если не принять мер по снижению нагрузок, то последствия выразятся в развитии профессионального или производственно-обусловленного заболевания в связи со снижением сопротивляемости организма.

Суммарное время регламентированных перерывов при работе с ПК зависит от категории трудовой деятельности и уровня нагрузки за рабочую смену. В таблице приведено суммарное время отдыха для каждой категории работ.

Таблица 24 – Суммарное время перерывов в зависимости от категории работы и нагрузки

Категория работы с ПЭВМ	Уровень нагрузки за рабочую смену при видах работ с ПЭВМ			Суммарное время регламентированных перерывов при 8-часовой смене, мин
	Группа А, количество знаков	Группа Б, количество знаков	Группа В, часов	
I	До 20000	До 15000	До 2	50
II	До 40000	До 30000	До 4	70
III	До 60000	До 40000	До 6	90

При разработке программы уровень нагрузки относился к группе В, категория работы III. Согласно таблице, суммарное время перерывов необходимо установить не менее 90 минут. По типовой инструкции по охране труда при работе на персональном компьютере ТОИ Р-45-084-01 [15] для данной категории работ требуется установить перерывы по 15 минут каждый трудовой час.

5.2.5 Перенапряжение зрительных анализаторов

При работе за ПК оператор читает текст в прямом свете, его глаза смотрят на источник света. При работе за дисплеем возникает следующая проблема: экран имеет высокую фоновую яркость, при ее понижении уменьшается контрастность изображения, что создает проблему для считывания информации. Поэтому пользователю приходится повышать яркость, что приводит к увеличению интенсивности вредных излучений и к утомлению глаз.

В настоящее время при работе за ПК визуальные параметры мониторов оказывают большую нагрузку на зрительные анализаторы, а именно на мышцы глаз. В результате чего мозгу сложно идентифицировать изображение.

В связи с этим параметры безопасности компьютерной техники указаны в соответствующих нормативных документах. После длительной нагрузки необходим период восстановления и отдыха. Если полного восстановления возможностей зрительного аппарата не происходит, то в детском и юношеском возрасте зрительная усталость приводит к нарушениям в аккомодационном механизме глаза, а в дальнейшем – к близорукости. В зрелом возрасте создаются предпосылки к развитию хронических глазных заболеваний

Повышенный уровень статического электричества в помещениях, оборудованных ПЭВМ, токи статического электричества чаще всего возникают при прикосновении персонала к любому из элементов ПЭВМ. Такие разряды опасности для человека не представляют, однако кроме неприятных ощущений могут привести к выходу оборудования из строя.

Для предотвращения образования и защиты от статического электричества в помещении используются нейтрализаторы и увлажнители, а полы имеют антистатическое покрытие в виде поливинилхлоридного антистатического линолеума.

Также в СанПиН 2.2.2/2.4.1340-03 [12] установлен максимальный допустимый электростатический потенциал экрана видеомонитора – 500 В.

В качестве мер уменьшения влияния вредных факторов на пользователя используются защитные фильтры для мониторов, увлажнители воздуха. Должны использоваться розетки с заземлением. Требуется проводить регулярную влажную уборку.

Необходимо заземление металлических электропроводных элементов оборудования, установка нейтрализаторов статического электричества (индукционных, высоковольтных, жидких и др.), увеличения поверхностей и объемной проводимости диэлектриков, что приводит к уменьшению

генерации электростатических зарядов или их отвод с наэлектризованного материала.

Для эффективной защиты также нужно обеспечить увеличение относительной влажности воздуха до (65 – 75) %, когда это возможно по условиям технологического процесса. Антистатическую обувь, антистатический халат, заземляющие браслеты применяют обычно в качестве индивидуальной защиты.

Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека. К опасностям использования электрического тока относятся возможность поражения электрическим током, а также воспламенения электронных устройств из-за воздействия различных условий – попадания влаги или нарушения изоляции. Поражение электрическим током может привести к ожогам, судорогам, повреждению нервной системы, а также смерти. Возникновение пожара может привести к последствиям, описанным в ГОСТ 12.1.033-81.

Во избежание негативных эффектов необходимо соблюдать правила пожарной и электрической безопасности. Подготовка к возникновению данных ситуаций должна производиться до начала работы.

Требования безопасности при эксплуатации электрооборудования регламентируются следующими нормативными актами:

- Правилами устройства электроустановок (издание шестое с отдельными разделами и главами в издании седьмом), утвержденными Главтехуправлением, Госэнергонадзором Минэнерго СССР 05.10.19 г.
- Правилами технической эксплуатации электроустановок потребителей, утверждёнными Приказом Минэнерго России от 13.01.2003 г. №6;
- Межотраслевыми правилами охраны труда (правилами безопасности) при эксплуатации электроустановок (ПОТ РМ 016-2001), утвержденными Постановлением Минтруда России от 05.01.2001 г. №3.

Согласно им:

- Электрооборудование, имеющее контакты для подключения заземления, должно быть заземлено, а помещения, где размещаются рабочие места с ПЭВМ, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации оборудования;
- Все крышки и защитные панели должны находиться на своих местах (при отсутствии крышки или защитной панели эксплуатация электрооборудования не допускается);
- При работе с электрооборудованием не допускать попадания влаги на поверхность электрооборудования, а также запрещается работать на электрооборудовании влажными руками;
- Вентиляционные отверстия электрооборудования не должны быть перекрыты находящимися вплотную стенами, мебелью, посторонними предметами;
- Выдергивание штепсельной вилки электроприбора необходимо осуществлять за корпус штепсельной вилки, при необходимости придерживая другой рукой корпус штепсельной розетки;
- подключение и отключение разъемов компьютеров и оргтехники должно производиться при отключенном питании (за исключением подключения и отключения USB-устройств);
- Удаление пыли с электрооборудования должно производиться в отключенном от электрической цепи состоянии
- Перед использованием электроприборов необходимо проверить надёжность крепления электророзетки, свериться с номиналом используемого напряжения;
- Корпуса штепсельных розеток и выключателей не должны содержать трещин, оплавлений и других дефектов, способных снизить защитные свойства или нарушить надёжность контакта;
- Кабели (шнуры) электропитания не должны содержать повреждений изоляции, сильных изгибов и скручиваний;

Короткое замыкание – это соединение двух точек с разным потенциалом с последующим увеличением тока и выделением большого количества тепла. Вследствие чего короткое замыкание может стать причиной пожара в помещении, при коротком замыкании от электрического тока могут пострадать люди, находящиеся в непосредственной близости от источника возникновения.

На рабочем месте короткое замыкание может быть вызвано либо неисправностью в проводке, либо при работе с компьютером/паяльным оборудованием, когда внутри корпуса создается разность фаз и ток может так же повредить всю электросеть.

Для защиты электрической сети от короткого замыкания предусмотрены устройства защитного отключения (УЗО), оснащенные устройствами автоматического отключения – автоматами и предохранителями. Кроме того, в помещении установлены датчики дыма, которые при возникновении возгорания, вызванного коротким замыканием, оповещают все здание о начавшемся пожаре. Таким образом, рабочее место полностью защищено от возможного короткого замыкания.

5.3 Экологическая безопасность

Разработка и эксплуатация системы распознавания почерка по клавиатурному почерку происходит в офисном помещении. Офис является источником следующих видов отходов:

- Твердые отходы: бумага, канцелярские принадлежности, комплектующие;
- Жидкие отходы: сточные воды;
- Люминесцентные лампы.

При использовании компьютера и другой вычислительной техники неизбежно устаревание моделей. В результате этого возникает необходимость замены и обновления оборудования. Чтобы уменьшить негативное воздействие на экологическую обстановку, утилизировать электронику нужно наиболее безопасным для окружающей среды способом, то есть обратиться в

специализированную компанию по утилизации. Такие компании действуют на всей территории Российской Федерации, в том числе и в Томской области.

Бумага может быть переработана и использована в качестве вторсырья. Для сбора макулатуры в России существуют специальные пункты приема. Некоторые из них предоставляют услугу вывоза макулатуры.

Сточные воды – один из источников загрязнения природных водоемов, так как они содержат различные загрязнения, в том числе мусор и примеси. В результате деятельности офисного предприятия происходит образование сточных вод. Для последующей очистки в системах канализации применяются отстаивание и фильтрация. Возможна дополнительная очистка с использованием озонаторов и ультрафиолета.

Отдельного внимания заслуживает вопрос утилизации люминесцентных ламп. Они покрыты люминесцентным веществом, имеют стеклянную оболочку и электроды. Внутри таких ламп находится инертный газ с парами ртути. В случае повреждения корпуса лампы, пары ртути попадают в атмосферу. Поэтому после окончания срока службы люминесцентные лампы необходимо сдавать на специальные предприятия по утилизации, имеющие специальную лицензию на данный вид деятельности.

В Томске к таким предприятиям относятся Экотом, Полигон, Утилизация. Необходимо отметить, что в целом при работе с компьютером существенного загрязнения окружающей среды не происходит и вредные выбросы не сравнимы с производственными.

5.4 Безопасность в чрезвычайных ситуациях

Наиболее вероятная чрезвычайная ситуация которая может возникнуть при работе с ПЭВМ – пожар, так как в современных ЭВМ очень высокая плотность размещения элементов электронных схем, в непосредственной близости друг от друга располагаются соединительные провода и кабели, при протекании по ним электрического тока выделяется значительное количество теплоты, при этом возможно оплавление изоляции и возникновение возгорания []. Возникновение других видов ЧС – маловероятно.

Причинами пожаров могут быть:

1. Несоблюдение правил пожарной безопасности;
2. Неисправность электропроводки;
3. Возгорание электроприборов по причине неисправности или в результате неправильной эксплуатации;
4. Курение вне специально отведенных мест.

В качестве противопожарных мероприятий должны быть применены следующие меры:

1. В помещении должны находиться средства тушения пожара, средства связи;
2. Электрическая проводка электрооборудования и осветительных приборов должна быть исправна;
3. Все сотрудники должны знать место нахождения средств пожаротушения и уметь ими воспользоваться, средств связи и номера экстренных служб.

В связи с возможностью возникновения пожара разработан следующий план действий:

1. В случае возникновения пожара сообщить о нем руководителю, постараться устранить очаг возгорания имеющимися силами при помощи первичных средств пожаротушения (огнетушитель порошковый, углекислотный О-1П0(з)-АВСЕ);
2. Привести в действие ручной извещатель, если очаг возгорания потушить не удастся;
3. Сообщить о возгорании в службу пожарной охраны по телефону 01 или 101, сообщить адрес, место и причину возникновения пожара;
4. Принять меры по эвакуации людей и материальных ценностей;
5. Встретить пожарную охрану, при необходимости сообщить всю необходимую информацию и оказать помощь при выборе наилучшего подхода к очагу возгорания;

Каждый сотрудник должен быть ознакомлен с планом эвакуации из помещения, с расположением выходов из здания. Плановые учебные эвакуации необходимы, чтобы подготовить сотрудников к действиям в чрезвычайной ситуации.

5.5 Вывод по разделу

В ходе выполнения работы над разделом «Социальная ответственность» были выявлены опасные и вредные факторы, воздействию которых может подвергнуться человек, занимающийся разработкой программы по распознаванию пользователей на основе параметров клавиатурного почерка.

Было установлено, что рабочее место удовлетворяет требованиям безопасности.

При работе в офисном помещении производятся отходы: бумага, канцелярские принадлежности, люминесцентные лампы и т.д. При надлежащей утилизации этих отходов (с помощью специальных фирм, имеющих лицензию на осуществление утилизации) загрязнение окружающей среды относительно мало.

Была рассмотрена наиболее распространенная чрезвычайная ситуация – пожар. Пожар может быть следствием короткого замыкания или неверной эксплуатации электроприборов, а также несоблюдения техники безопасности. Поэтому сотрудники офиса проходят инструктаж по правилам пожарной безопасности, проходят учебные эвакуации.

Правовые и организационные вопросы обеспечения безопасности регулируются государственными органами. Правила и нормы для обеспечения нормальных условий труда всех сотрудников устанавливаются на государственном уровне.

Заключение

В ходе исследования были получены новые теоретические сведения и особенности клавиатурного почерка, которые были использованы в качестве критериев идентификации пользователей информационной системы.

Было поставлено несколько задач: задача Data Mining – сбор данных, идентификация и установление чувствительности методов.

Задача Data Mining была решена в два этапа. В первом были собраны реальные сессии пользователей. Во втором этапе, из-за малого объема собранных сессий, были смоделированы дополнительные активности пользователей.

Для моделирования активностей пользователей на основе реальных сессий было реализовано программное приложение.

Решение задачи идентификации сводилось к изучению различных методов идентификации. В ходе работы были рассмотрены существующие подходы, а также предложены свои решения.

Результаты идентификации были оценены, исходя из ошибок I и II, полученных при распознавании пользователей. Чувствительность методов была установлена в соответствии с поставленной задачей – чем меньше незаконных пользователей проникнет к системе, тем она устойчивее.

Для идентификации пользователей, а также для оценки ошибок I и II рода разработано программное приложение.

Для повышения эффективности системы, в дальнейшем, планируется изучить другие методы идентификации, а также рассмотреть возможность комбинирования нескольких методов на основе различных биометрических данных человека.

Conclusion

During the study, new theoretical knowledge and different features of keyboard writing were obtained. They were used as criteria to identify users of the information system.

Several tasks were set during the thesis: Data Mining problem – collecting of the data, the user identification problem and establishment of the sensitivity of methods problem.

The Data Mining problem was solved through two steps. The first step involved the collection of data in real user sessions. In the second step, additional user activities were modelled due to the small number of collected sessions.

A software application was developed for modelling user activities based on the real sessions.

The solution of the user identification problem led to the study of various user identification methods. Throughout the work, existing approaches were considered and own solutions were proposed as well.

The results of user identification were estimated by the values of I and II errors, which were received during user recognition. The sensitivity of the methods was established according to the applied problem – the less illegal users that could enter into the system, the more stable it was.

A software application was developed to identify users in the information system. The software estimates the values of errors I and II as well.

Other identification methods are planned to be explored to improve the system's efficiency in the sequel. Moreover, it is planned to consider the possibility of combining several methods based on different biometric data.

Список литературы

1. Клавиатурный почерк как средство аутентификации // SecurityLab.ru
URL: <https://www.securitylab.ru/blog/personal/aguryanov/29985.php> (дата обращения: 15.05.2020).
2. Глобальное исследование утечек конфиденциальной информации в первом полугодии 2019 года // InfoWatch URL: https://www.infowatch.ru/sites/default/files/report/analytics/russ/Global_Data_Leaks_Report_2019_half_year.pdf (дата обращения: 15.05.2020).
3. Кочегурова Е.А., Мартынова Ю.А. Особенности непрерывной идентификации пользователей на основе свободных текстов в режим скрытого мониторинга // Программирование. - 2020. - №1. - С. 15-28.
4. Ворона В.А., Тихонов В.А. Системы контроля и управления доступом. – М.: Горячая линия–Телеком, 2013. – 272 с.: ил.
5. Васильев В.И., Жумажанова С.Г., Ложников П.С., Сулавко А.Е. Оценка идентификационных возможностей биометрических признаков от стандартного периферийного оборудования // Вопросы защиты информации. 2016. №1. С. 12-20.
6. Назаров В.Е., Карасева Г.Т., Успенский Ю.П., Джагацпанян И.Э. Оценка риска патологических состояний с помощью анализа газового состава выдыхаемого воздуха // Новые информационные технологии в автоматизированных системах. - 2013. - №4. - С. 218-225.
7. Ершов К.С., Романова Т.Н. Анализ и классификация алгоритмов кластеризации // Новые информационные технологии в автоматизированных системах. - 2016. - №19. - С. 274-279.
8. Стадников Д.А. Обзор технологий интеллектуального анализа данных // Молодежный вестник ИРГТУ. - 2017. - №4. - С. 2.
9. Pisani P.H., Lorena A.C. A systematic review on keystroke dynamics // Journal of the Brazilian Computer Society. - 2013. - №19. - С. 573–587.

10. Акаткин Ю.М., Конявский В.А., Кристальный Б.В. Вопросы информационной безопасности при оказании услуг населению в электронной форме // Информационные ресурсы России. - 2014. - №1. - С. 15-22.

11. Глухов Д.О., Глухова Т.М. Генераторы случайных чисел для расчета надежности строительных конструкций // Вестник Полоцкого Государственного Университета. - 2008. - №3. - С. 60-68.

12. СанПиН 2.2.2/2.4.1340-03 "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы" (с изменениями на 21 июня 2016 года) // Техноэсперт URL: <http://docs.cntd.ru/document/901865498> (дата обращения: 01.05.2020).

13. СанПиН 2.2.4.3359-16 "Санитарно-эпидемиологические требования к физическим факторам на рабочих местах" // Техноэсперт URL: <http://docs.cntd.ru/document/420362948> (дата обращения: 01.05.2020).

14. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений // Техноэсперт URL: <http://docs.cntd.ru/document/901704046> (дата обращения: 01.05.2020).

15. ТОИ Р-45-084-01 Типовая инструкция по охране труда при работе на персональном компьютере // Техноэсперт URL: <http://docs.cntd.ru/document/1200030047/> (дата обращения: 01.05.2020).

16. Чрезвычайные ситуации при работе с ПЭВМ // Студопедия URL: https://studopedia.ru/8_107307_osveshchenie-pomeshcheniy-vichislitelnih-tsentrov.html (дата обращения: 01.05.2020).

Приложение А

(обязательное)

Таблица А.1 – Частотность букв русского алфавита в %

№	Буква	Частотность
1	О	10.97%
2	Е	8.45%
3	А	8.01%
4	И	7.35%
5	Н	6.70%
6	Т	6.26%
7	С	5.47%
8	Р	4.73%
9	В	4.54%
10	Л	4.40%
11	К	3.49%
12	М	3.21%
13	Д	2.98%
14	П	2.81%
15	У	2.62%
16	Я	2.01%
17	Ы	1.90%
18	Ь	1.74%
19	Г	1.70%
20	З	1.65%
21	Б	1.59%
22	Ч	1.44%
23	Й	1.21%
24	Х	0.97%
25	Ж	0.94%
26	Ш	0.73%
27	Ю	0.64%
28	Ц	0.48%
29	Щ	0.36%
30	Э	0.32%
31	Ф	0.26%
32	Ъ	0.04%
33	Ё	0.04%