

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

Школа: Информационных технологий и робототехники

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Отделение школы (НОЦ): Информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы

Идентификация пользователей компьютерных систем на основе динамических характеристик клавиатурного почерка

УДК 004.421.056.523:57.087

Студент

Группа	ФИО	Подпись	Дата
8ВМ6Г	Горохова Екатерина Сергеевна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Кочегурова Елена Алексеевна	к.т.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Рыжакина Татьяна Гавриловна	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ассистент	Авдеева Ирина Ивановна	-		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Кочегурова Елена Алексеевна	к.т.н.		

Томск – 2018 г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результатов	Результат обучения (выпускник должен быть готов)
	Общепрофессиональные компетенции
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять специальную лексику и профессиональную терминологию языка.
	Профессиональные компетенции
P5	Выполнять инновационные инженерные проекты по разработке аппаратных и программных средств автоматизированных систем различного назначения с использованием современных методов проектирования, систем автоматизированного проектирования, передового опыта разработки конкурентно способных изделий.
P6	Планировать и проводить теоретические и экспериментальные исследования в области проектирования аппаратных и программных средств автоматизированных систем с использованием новейших достижений науки и техники, передового отечественного и зарубежного опыта. Критически оценивать полученные данные и делать выводы.
P7	Осуществлять авторское сопровождение процессов проектирования, внедрения и эксплуатации аппаратных и программных средств автоматизированных систем различного назначения.
	Общекультурные компетенции
P8	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современного оборудования и приборов, в управлении коллективом.
P9	Осуществлять коммуникации в профессиональной среде и в обществе в

	целом, активно владеть иностранным языком, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности, в том числе на иностранном языке.
P10	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.
P11	Демонстрировать способность к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности, способность к педагогической деятельности.

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

Школа Информационных технологий и робототехники
Направление подготовки (специальность) 09.04.01 Информатика и вычислительная техника
Отделение школы (НОЦ) Информационных технологий

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

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
8ВМ6Г	Гороховой Екатерине Сергеевне

Тема работы:

Идентификация пользователей компьютерных систем на основе динамических характеристик клавиатурного почерка
--

Утверждена приказом директора (дата, номер)	
---	--

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	
---------------------------------	--

- | | |
|--|--|
| | <ol style="list-style-type: none">1. Существующие алгоритмы распознавания клавиатурного почерка;2. Требование реализовать клиент-серверную архитектуру;3. Требование организовать сбор характеристик клавиатурного почерка.4. Требование реализовать сравнение нескольких почерков на основе векторного критерия. |
|--|--|

Перечень подлежащих исследованию, проектированию и разработке вопросов	<ol style="list-style-type: none"> 1. Рассмотреть существующие алгоритмы анализа компьютерного почерка; 2. Определить функционал приложения; 3. Спроектировать архитектуру программного приложения; 4. Реализовать получение образца клавиатурного почерка пользователя; 5. Реализовать сравнение образца почерка пользователя эталонами; 6. Реализовать хранение результатов проверки в лог-файле.
---	---

Перечень графического материала	Презентация в формате *.pptx на 21 слайде
--	---

Консультанты по разделам выпускной квалификационной работы

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Рыжакина Татьяна Гавриловна
Социальная ответственность	Авдеева Ирина Ивановна

Названия разделов, которые должны быть написаны на русском и иностранном языках:

Обзор технологий и решений
Проектирование
Реализация и тестирование

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

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кочегурова Елена Алексеевна	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ6Г	Горохова Екатерина Сергеевна		

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



федеральное государственное автономное образовательное учреждение
высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Школа Информационных технологий и робототехники
Направление подготовки (специальность) 09.04.01 Информатика и вычислительная техника

Отделение школы (НОЦ) Информационных технологий

Форма представления работы:

магистерская диссертация

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы: 29.05.2018

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
28.05.2018	Основная часть	75
28.05.2018	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
28.05.2018	Социальная ответственность	10

Составил преподаватель:

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

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Кочегурова Е.А.	к.т.н.		

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

Студенту:

Группа	ФИО
8ВМ6Г	Гороховой Екатерине Сергеевне

Школа	ИШИТР	Отделение	Информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Работа с информацией, представленной в российских и иностранных научных публикациях, аналитических материалах, статических бюллетенях и изданиях, нормативно-правовых документах; анкетирование; опрос.
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	Проведение предпроектного анализа. Определение целевого рынка и проведение его сегментирования. Выполнение SWOT-анализа проекта
2. Определение возможных альтернатив проведения научных исследований	Определение целей и ожиданий, требований проекта. Определение заинтересованных сторон и их ожиданий.
3. Планирование процесса управления НИ: структура и график проведения, бюджет, риски и организация закупок	Составление календарного плана проекта. Определение бюджета НИ
4. Определение ресурсной, финансовой, экономической эффективности	Проведение оценки экономической эффективности разработки системы распознавания пользователя по клавиатурному почерку

Перечень графического материала (с точным указанием обязательных чертежей):

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

Дата выдачи задания для раздела по линейному графику	14.03.2018
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Рыжакина Татьяна Гавриловна	Кандидат экономических наук		14.03.2018

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ6Г	Горохова Екатерина Сергеевна		14.03.2018

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

Группа	ФИО
8ВМ6Г	Гороховой Екатерине Сергеевне

Школа	ИШИТР	Отделение	Информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения

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 по распознаванию. 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.

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Производственная безопасность

Возможные вредные факторы в офисном помещении:

- Недостаточная вентиляция;
- Недостаточное или неправильное освещение;
- Пониженная или повышенная температура воздуха;
- Повышенный уровень электромагнитного излучения;
- Нервно-психические перегрузки.

Возможные опасные факторы в офисном помещении:

- Короткое замыкание;
- Электрический ток.

2. Экологическая безопасность

В процессе разработки и эксплуатации системы идентификации возможно образование следующих видов отходов:

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

3. Безопасность в чрезвычайных ситуациях	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.
4. Правовые и организационные вопросы обеспечения безопасности	Параметры рабочего места офисного работника регулируются ГОСТ 12.2.032–78 ССБТ, СанПиН 2.2.2/2.4.1340–03, ФЗ-197.

Дата выдачи задания для раздела по линейному графику	14.03.2018
---	-------------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ассистент	Авдеева Ирина Ивановна	-		14.03.2018

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ6Г	Горохова Екатерина Сергеевна		14.03.2018

РЕФЕРАТ

Выпускная квалификационная работа состоит из 109 с., 33 рис., 30 табл., 37 источников, 3 прил.

Ключевые слова: клавиатурный почерк, биометрия, аутентификация, информационная безопасность, идентификация пользователя.

Объектом исследования является разрабатываемая система распознавания клавиатурного почерка.

Цель работы – развитие задач теории распознавания, формирование новых динамических характеристик клавиатурного почерка (КП) и разработка на этой основе способов и алгоритмов идентификации и аутентификации пользователя с реализацией прототипа программного приложения.

В процессе исследования проводились работы по исследованию существующих алгоритмов и методик анализа клавиатурного почерка. В результате работы был создан алгоритм, а также спроектирована и реализована система идентификации пользователей компьютерных сетей на основе динамических характеристик клавиатурного почерка.

Основные конструктивные, технологические и технико-эксплуатационные характеристики: архитектура типа клиент-сервер.

Степень внедрения: в настоящий момент система внедрена в работу ООО «ЭДИСОН» и используется в тестовом режиме. Область применения: аутентификация пользователей по клавиатурному почерку, скрытый мониторинг клавиатурного почерка с целью определения подмены оператора.

Экономическая эффективность/значимость работы обусловлена стоимостью информации, подлежащей защите с помощью разрабатываемого средства аутентификации. В будущем планируется дальнейшая разработка и совершенствование системы в целях повышения точности работы, отказоустойчивости и защищенности.

Определения, обозначения, сокращения, нормативные ссылки

КП – клавиатурный почерк

ВУК – время удержания клавиши

ЭВМ – электронная вычислительная машина

UML – Unified Modeling Language

TCP – transmission control protocol

JSON – JavaScript Object Notation

OS – Operating System

FRR – False Rejection Rate

FAR – False Acceptance Rate

Оглавление

РЕФЕРАТ	11
Определения, обозначения, сокращения, нормативные ссылки	12
Введение	16
1 Обзор технологий и решений.....	17
1.1 Цели и задачи.....	17
1.2 Современное состояние проблемы.....	18
1.2.1 Характеристики клавиатурного почерка	18
1.2.2. Сбор данных	20
1.2.3 Алгоритмы распознавания на основе клавиатурного почерка.....	21
1.2.4 Показатели эффективности алгоритмов	26
1.3 Прикладные задачи	28
1.4 Актуальность исследования.....	29
2 Проектирование	31
2.1 Функциональные возможности системы.....	31
2.2 Алгоритм работы программы.....	32
2.3 Выбор средств разработки	40
2.4 Архитектура системы	42
3 Реализация и тестирование.....	45
3.1 Описание интерфейса	45
3.2 Сбор тестовых данных	47
3.3 Тестирование программы	50
4 Финансовый менеджмент	58
4.1 Предпроектный анализ	58
4.1.1 Потенциальные потребители результатов исследования	58
4.1.2 Анализ конкурентных технических решений с позиции.....	59
ресурсоэффективности и ресурсосбережения	59
4.1.3 SWOT-анализ.....	60
4.1.4 Оценка готовности проекта к коммерциализации.....	61

4.1.5 Методы коммерциализации результатов научно-технического исследования.....	61
4.2 Инициация проекта	62
4.2.1 Ограничения и допущения проекта.	63
4.3 Планирование управления научно-техническим проектом.....	64
4.3.1 План проекта.....	64
4.3.2 Бюджет научного исследования	64
4.3.2 Организационная структура проекта	66
4.3.3 План управления коммуникациями проекта.....	67
4.3.4 Реестр рисков проекта	68
4.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования .	68
4.4.1 Оценка абсолютной эффективности исследования.....	68
4.4.2 Оценка сравнительной эффективности исследования	70
5 Социальная ответственность.....	74
5.1 Введение.....	74
5.2 Производственная безопасность.....	74
5.3 Экологическая безопасность.....	81
5.4 Безопасность в чрезвычайных ситуациях.....	82
5.5 Правовые и организационные вопросы обеспечения безопасности.....	83
5.6 Вывод.....	84
Заключение.....	85
Conclusion	87
Список литературы.....	88
Список публикаций	92
Приложение А.....	93
Приложение Б	100
1 Overview of technologies and solutions	101
1.1 Goals and objectives.....	101
1.2 Present state of the problem	102

1.3 Practical application	108
1.4 Relevance of the research.....	109
Приложение В.....	110

Введение

В современном мире важность информации сложно переоценить. Для её защиты широко применяется аутентификация и авторизация. Аутентификация – это процедура проверки подлинности. Под авторизацией понимается процедура предоставления пользователю определенных прав.

Различают три типа методов аутентификации:

- Парольные;
- Имущественные;
- Биометрические.

Наиболее распространены в настоящее время парольные методы. Они просты в реализации и обладают высокой точностью: всегда можно определить, совпадают пароли или нет. Имущественные методы предполагают использование специальных устройств, например, USB-ключей, для подтверждения личности.

Биометрические методы основаны на определении личности человека по присущим только ему признакам. Основное достоинство биометрических методов состоит в том, что такие признаки невозможно украсть или передать другому человеку.

Биометрические методы делятся на физиологические и поведенческие. К физиологическим методам относится сканирование отпечатков пальца, радужной оболочки глаза, формы лица и т.д. Поведенческие методы предполагают анализ речи, рукописного и клавиатурного почерка, походки. Поведенческие характеристики сложнее распознать с высокой точностью, но вместе с тем, сложнее и подделать.

Целью работы является разработка методики и программного обеспечения для идентификации пользователей компьютерных систем по динамическим характеристикам клавиатурного почерка.

1 Обзор технологий и решений

1.1 Цели и задачи

Цель работы: Развитие задач теории распознавания, формирование новых динамических характеристик клавиатурного почерка (КП) и разработка на этой основе способов и алгоритмов идентификации и аутентификации пользователя с реализацией прототипа программного приложения.

Для достижения поставленной цели необходимо решить следующие задачи.

1. Анализ тенденций и подходов к проблематике использования биометрических параметров в целях аутентификации пользователя.

2. Исследование современных подходов к распознаванию КП пользователей.

3. Анализ существующих характеристик КП. Анализ их идентификационной способности для распознавания пользователей по существующим характеристикам. Создание векторного динамического показателя КП для сравнения двух и более образцов клавиатурного почерка.

4. Анализ механизма формирования характеристик КП при нажатии клавиш. Получение математической модели формирования характеристик КП.

5. Разработка алгоритма сбора образцов КП для получения векторного динамического показателя. Создание программного приложения для сбора статистики.

6. Сбор статистики показателей КП пользователей. Создание банка данных показателей КП для разных категорий населения.

7. Анализ статистических данных, обработка данных, выделение закономерностей. Исследование разрешающей способности статистических данных.

8. Исследование существующих критериев различия (близости) характеристик КП. Создание векторного дискриминантного критерия оценки совпадения образцов КП с использованием частотности букв русского языка и существующих метрик близости.

9. Разработка гибридного метода анализа КП для выявления попытки несанкционированного входа и аутентификация зарегистрированного пользователя.

10. Разработка программного приложения для получения образцов КП пользователей, аутентификации зарегистрированных пользователей и предотвращения попыток несанкционированного доступа.

1.2 Современное состояние проблемы

1.2.1 Характеристики клавиатурного почерка

В [1] и [2] приводится характеристика стабильности почерка пользователей. Оценка означает пригодность почерка для анализа клавиатурного почерка. Предлагается отсеивать пользователей, чей почерк недостаточно стабилен, во избежание дискредитации системы.

Таблица 1 – Характеристики стабильности клавиатурного почерка пользователей

Ошибки , %	Аритмичность , %	Скорость , зн./мин	Характеристика перекрытий		Оценка
			Число перекрытий , %	Используемо е число пальцев	
Менее 2	Менее 10	Более 200	Более 50	Все	Отлично
Менее 4	Менее 15	Более 150	Более 30	Большинство	Хорошо
Менее 8	Менее 20	Более 100	Более 10	Несколько	Удовл.
Более 8	Более 20	Менее 100	Менее 10	По одному	Неуд.

Как правило, необходимо 6 месяцев работы с клавиатурой, для того чтобы клавиатурный почерк стабилизировался.

Клавиатурный почерк характеризуется рядом параметров:

- Время удержания клавиши (ВУК);
- Паузы между нажатиями;
- Наличие наложений;
- Число ошибок при вводе;
- Степень аритмичности при наборе;
- Скорость набора;
- Особенности использования служебных клавиш.

На рисунке 1 показаны основные характеристики КП: время удержания клавиши и пауза между нажатиями.

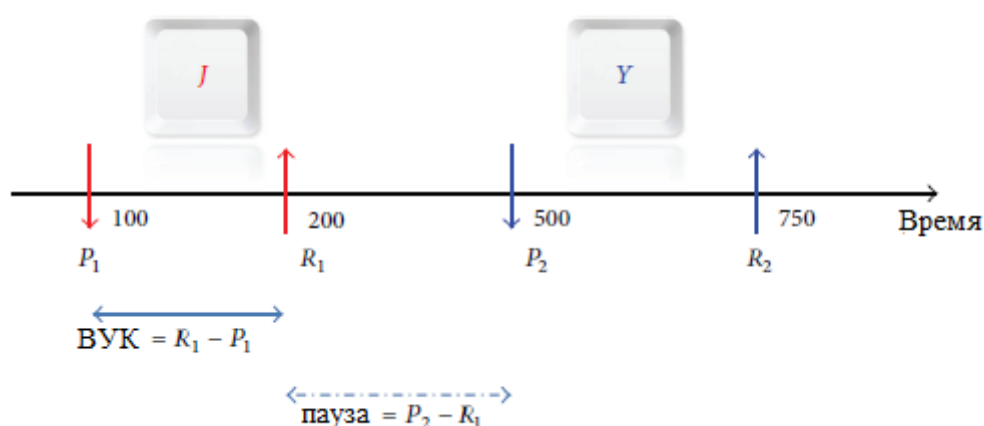


Рисунок 1 – ВУК и пауза

Большая часть исследователей КП анализируют в своих работах [3-7] 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 2 приведены эти характеристики.

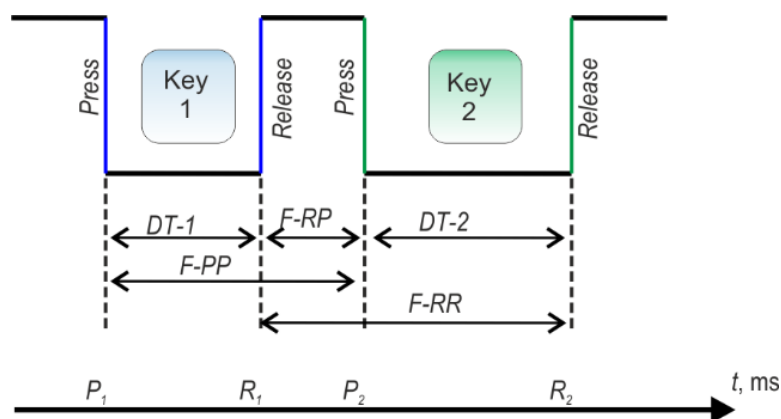


Рисунок 2 – Диграфы

Основные сложности при анализе клавиатурного почерка:

- Разброс параметров клавиатурного почерка в зависимости от психо-физического состояния пользователя;
- Разброс параметров клавиатурного почерка в зависимости от используемой клавиатуры. В работе [1] говорится, что исследование КП пользователя на разных клавиатурах дало разброс вероятности аутентификации на 0.5%.
- Необходимость сбора большого количества статистических данных для каждого исследования КП, отсутствие готовых баз данных с образцами КП.

1.2.2. Сбор данных

При непрерывном сборе данных необходимо возникает ряд дополнительных вопросов:

- Объем данных для формирования шаблонов. Условно объемы данных при непрерывном исследовании клавиатурной динамики принято разделять на небольшой (менее 1000 нажатий), средний (менее 6000 нажатий) и большой (более 6000 нажатий) на каждого пользователя. Иногда в свободных текстах анализируется количество слов. По данным [8] 57% исследователей анализируют короткие тексты, 24% - длинные, оставшиеся 19% распределены между текстами, содержащие только цифры и тексты неизвестного содержания.

- Формирование и использование общедоступных БД профилей пользователей. Первые БД клавиатурных нажатий появились в 2009 года. БД содержат разное количество пользователей (30-300) и разное число повторений в сессию. Но абсолютное большинство БД содержат клавиатурные данные, собранные на основании паролей и фиксированных текстов и, следовательно, малопригодны для скрытого и непрерывного мониторинга. Однако главной проблемой здесь является полное отсутствие подобных ресурсов с шаблонами пользователей на русском языке. А это приводит к необходимости сбора данных каждым исследователем. И соответственно затрудняет сопоставление результатов в равных условиях.

1.2.3 Алгоритмы распознавания на основе клавиатурного почерка

Все алгоритмы распознавания пользователя по клавиатурному почерку можно разделить на те, где идентификация происходит по набору произвольного или фиксированного (заранее заданного) текста, например, короткого пароля. Алгоритмы первого типа позволяют отслеживать несанкционированную подмену оператора, в то время как алгоритмы второго типа проще в реализации.

В любом случае система включает два режима работы: обучение и идентификация.

В процессе аутентификации необходимо решить задачу классификации пользователя на своего и чужого. При этом используются [2-15]:

- Геометрические методы, рассчитывающие меру близости вектора к другому вектору (мера Хэмминга, Евклидова мера, расстояние Махаланобиса и т.д.);

- Методы на основе ИНС;
- Методы параметрического обучения классификатора.

Метод параметрического обучения предполагает для каждого образца клавиатурного почерка расчет следующих величин.

Аритмичность набора α и математическое ожидание m_{pause}

$$\bar{m}^{\text{pause}} = \frac{\sum_{i=1}^{n-1} t_i^{\text{pause}}}{(n-1)}; \alpha = \sqrt{\frac{\sum_{i=1}^{n-1} \left(\frac{t_i^{\text{pause}}}{t_{\text{max}}^{\text{pause}}} - \bar{m}^{\text{pause}} \right)^2}{n-2}}. \quad (1)$$

Аритмичность времени удержания клавиш β и математическое ожидание m -press

$$\bar{m}^{\text{press}} = \frac{\sum_{i=1}^n t_i^{\text{press}}}{n}; \beta = \sqrt{\frac{\sum_{i=1}^n \left(\frac{t_i^{\text{press}}}{t_{\text{max}}^{\text{press}}} - \bar{m}^{\text{press}} \right)^2}{n-1}}. \quad (2)$$

Скорость набора

$$s = \frac{(t_n^{\text{up}} - t_1^{\text{down}})}{60} s_{\text{max}}, \quad (3)$$

Где s_{max} – максимальная скорость набора текста.

Полученные результаты для метода параметрического классификатора: FRR = 8%, FAR=7%.

При использовании нейронной сети для определения типа личности (экстраверт интроверт) была получена максимальная точность в 95%. В данной работе использовался микрофон для записи звуков клавиатурных нажатий. При этом анализировалась интенсивность ударов по клавишам.

При использовании для аутентификации пользователей по клавиатурному почерку Многослойного персептрона полученные результаты включают FAR=0% и FRR=1%.

Также перспективным методом является использование трехслойного персептрона Розенблатта.

- Первичный слой – входной. Он состоит из k формальных нейронов с линейной активаторной функцией, где k – размерность входного вектора с параметрами клавиатурного почерка.

- Второй слой является скрытым. Он состоит из k формальных нейронов сигмоидной активаторной функцией.

- Третий слой – выходной. Он включает из p формальных нейронов сигмоидной активаторной функцией, где p – число зарегистрированных в системе пользователей.

Для повышения точности алгоритмов распознавания клавиатурного почерка в работе [11] предлагается использовать разделение ВУК на нажатия с наложениями и без наложений. В качестве обоснования можно привести график биполярного нормального распределения ВУК.

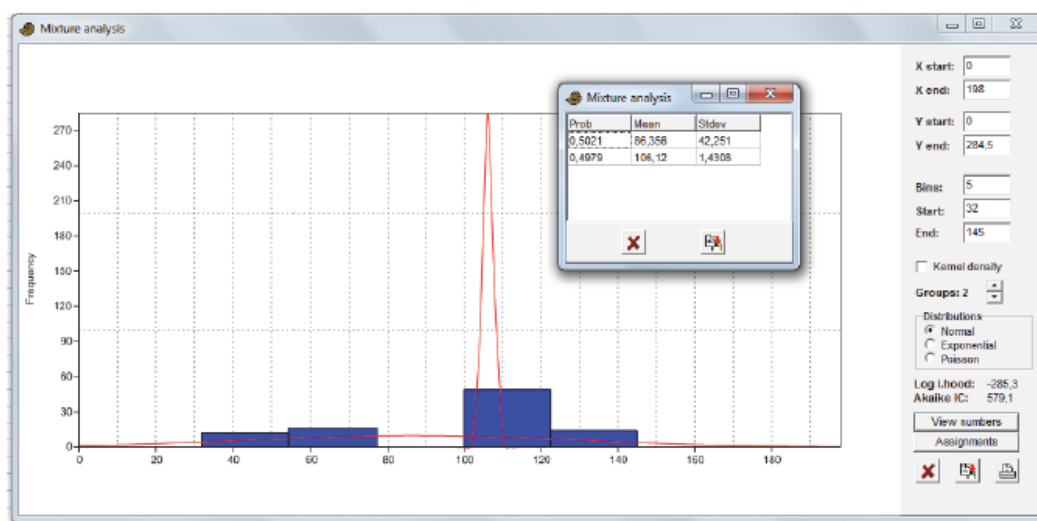


Рисунок 3 – График биполярного нормального распределения ВУК

В ход эксперимента авторами было зафиксировано время удержания клавиши «А» 91 раз:

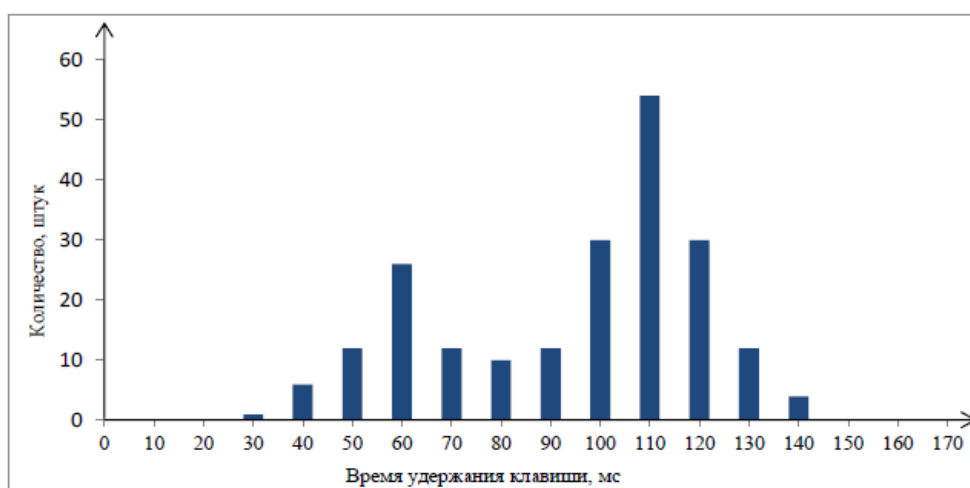


Рисунок 4 – Бимодальное распределение ВУК

Для определения расстояния между векторами клавиатурного почерка возможно использование Евклидовой меры, меры Махаланобиса, Манхеттонской дистанции. В качестве классификаторов использовались:

- Метод К-средних;
- К-ближайших соседей;
- Байесовский классификатор;
- Нечеткая логика (Fuzzy logic);
- Нейронные сети;
- SVM – метод опорных векторов.

Также использовалась мера Хэмминга для сравнения двоичных векторов. Дистанция Хэмминга определяет суммарное число различий между элементами векторов, находящихся на одной позиции. Двоичный вектор получается путем сравнения тестового образца и эталона клавиатурного почерка. Кроме того, в данной работе предлагается использовать триграф – три набранных подряд символа. Длительность триграфа – это время между нажатием первой и третьей клавиши. В результате наилучший полученный результат распознавания включает FAR = 0.04% и FRR=6.4%.

Метод многосвязного представления событий при пользовании клавиатурой может использоваться для анализа произвольного текста. Схематичное изображение механизма разбиения на классы представлено на рисунке.

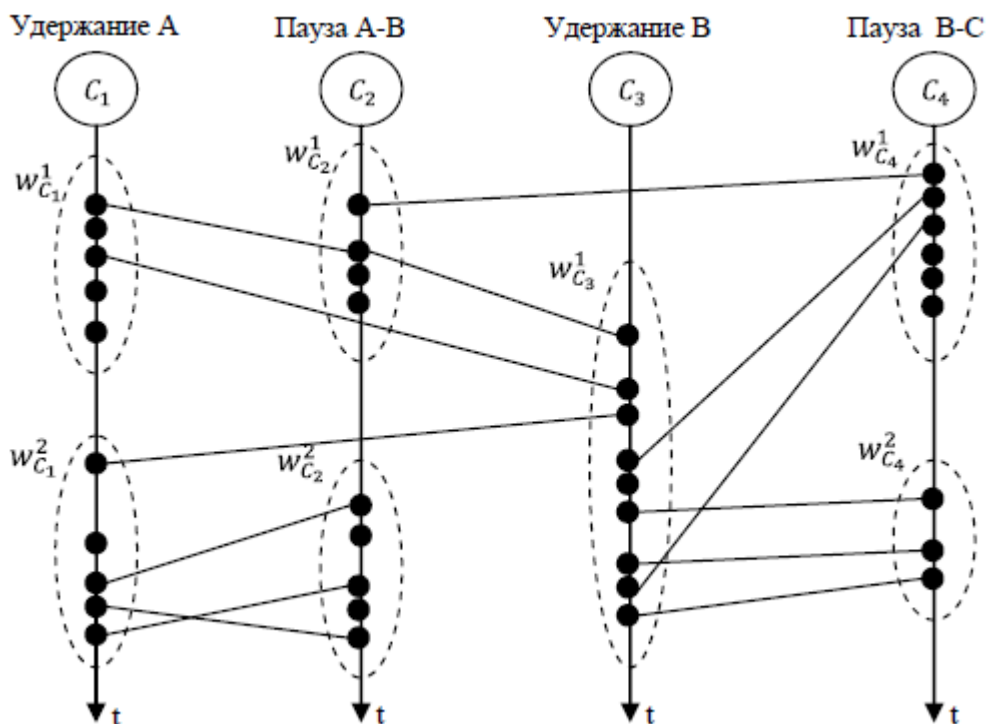


Рисунок 5 – Многосвязное представление клавиатурного почерка

В ходе экспериментов пользователи вводили текст длиной 10000 знаков. Была успешно определена подмена пользователя.

В ряде случаев задача аутентификации по клавиатурному почерку решается путем исключения грубых ошибок и, затем, решения задачи проверки гипотезы о равенстве центров распределения двух нормальных генеральных совокупностей. Результаты актуальны для ввода коротких парольных фраз. Точность алгоритма достигает 97%.

Из рисунка 6 следует, что оценка расстояния (метрики) является самым популярным методом распознавания и достигает по частоте использования 23%. В рамках этого подхода вычисляется расстояние между эталонным и текущим профилем пользователя, которое затем сопоставляется с пороговым значением. Наиболее часто используемыми метрическими расстояниями являются меры Евклида, Манхэттенская, Хэмминга, Махаланобиса [7-19]. В том числе на основе относительных расстояний для длинного и свободного текста [20].

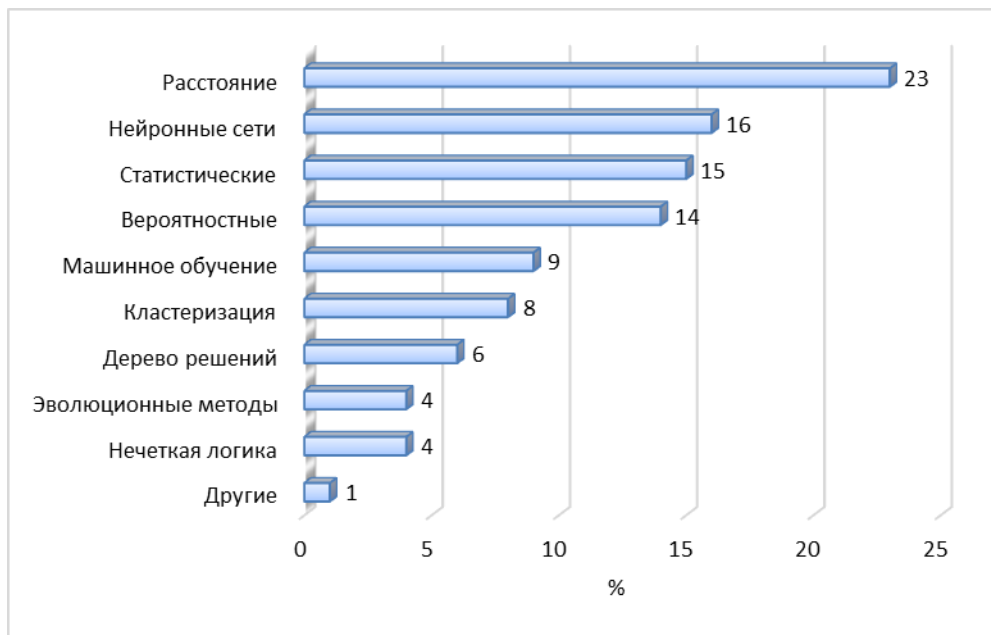


Рисунок 6 – Частота использования методов клавиатурного распознавания

Таким образом, можно сделать вывод, что тема распознавания пользователя по клавиатурному почерку в настоящее время является актуальной, ей посвящено множество работ. Полученные результаты делают возможным применение систем распознавания клавиатурного почерка на практике.

1.2.4 Показатели эффективности алгоритмов

Для анализа эффективности работы аутентификационных алгоритмов наиболее часто используются ошибки первого и второго рода, также называемые FRR и FAR.

- FRR – отказ в доступе;
- FAR – ложный доступ.

При сравнении образца клавиатурного почерка с эталоном возможны следующие варианты развития событий [9]:

- Образцы принадлежат одному и тому же пользователю, и система определяет образцы как схожие – это ожидаемое поведение системы;
- Образцы принадлежат разным пользователям, и система определяет их как не схожие – это тоже ожидаемое поведение системы;

- Образцы принадлежат одному и тому же пользователю, но система определяет их как не схожие – FRR, так как опровергается верная гипотеза;

- Образцы принадлежат разным пользователям, но система определяет их как схожие – FAR, так как принимается ложная гипотеза.

Гипотетически ошибки FRR и FAR варьируются в зависимости от уровня чувствительности алгоритма (порогового значения) как показано на рисунке 7а): когда одна ошибка уменьшается, другая увеличивается.

Более высокие значения FAR обычно предпочтительнее в системах, где безопасность не имеет первостепенной важности, тогда как более высокие значения FRR является предпочтительным в приложениях с высокой степенью защиты. Компромисс между FAR и FRR должен определяться целями конкретной прикладной задачи. Например, малый пропуск нелегальных пользователей (FAR) соответствует высокому пороговому значению (чувствительности), но приводит к большому отклонению зарегистрированных пользователей (FRR), т.е. к их низкому пропуску.

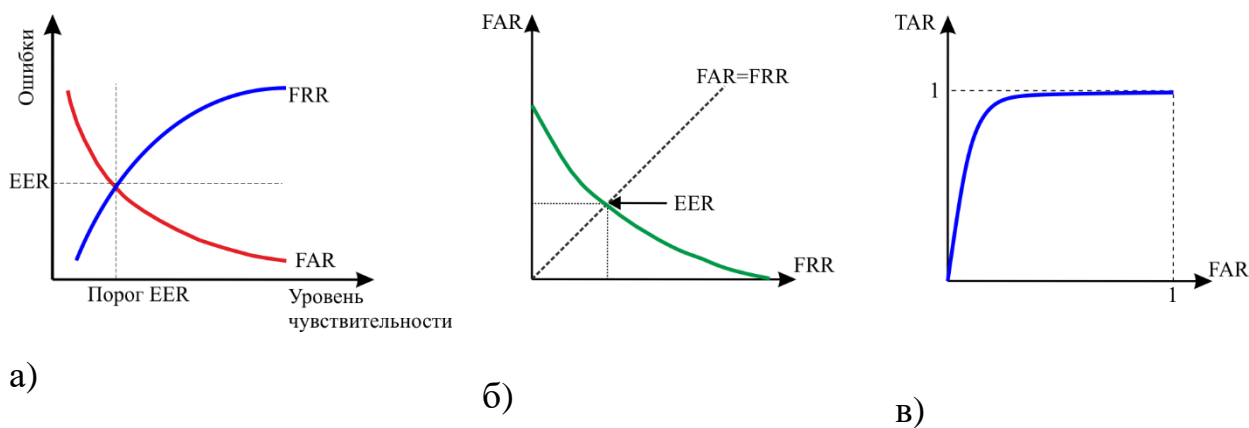


Рисунок 7 – Показатели эффективности клавиатурной идентификации

Между ошибками FAR и FRR существует компромисс, т. е. можно уменьшить одну погрешность за счет другой, регулируя порог принятия решения. На рисунке 7-б) представлена кривая DET (Detection Error Trade-off), позволяющая определить компромисс между ошибками I рода (FRR) и

II рода (FAR). По ней можно установить одну из желаемых ошибок и увидеть значение другой.

Часто используемая Equal Error Rate (EER) - равная частота ошибок - для определения общей точности системы распознавания. EER представляет значение ошибки, когда FAR и FRR принимают равные значения. В отличие от FAR и FRR, ошибка EER не зависит от уровня чувствительности алгоритма аутентификации. Чем ниже значение EER, тем эффективнее система распознавания при заданном пороговом значении. Независимый от порога EER более подходит для оценки эффективности алгоритмов распознавания.

Показатели FAR, FRR и EER весьма популярны и перспективны в системах клавиатурной аутентификации. Менее часто используемая характеристика ROC (Receiver Operating Characteristic) [21-23], позволяет увидеть предельные значения показателей эффективности. Характеристика ROC отображает компромисс между верным (TAR) и ошибочным (FAR) пропуском пользователя при различных пороговых значениях. Верхний левый угол графика представляет собой идеальную точку, где TAR равно единице, а FAR равно нулю.

1.3 Прикладные задачи

В данной работе внимание сфокусировано на исследовании клавиатурного почерка и создании методики анализа клавиатурного почерка для аутентификации пользователя.

С помощью анализа клавиатурного почерка возможно решить ряд прикладных задач, а именно:

- Своевременное определение подмены пользователя.

Предположим, аутентифицированный пользователь покинул рабочее место, оставив компьютер незаблокированным. В это время злоумышленник может получить доступ ко всем данным аутентифицированного пользователя. Постоянный скрытый мониторинг клавиатурного почерка позволяет

определять подмену пользователя в любой момент работы с компьютером, даже если аутентификация уже была успешно пройдена.

- Клавиатурный почерк может использоваться как дополнительное средство аутентификации. Для этого при вводе логина и пароля необходимо также производить анализ динамических характеристик работы на клавиатуре.

- Идентификация психо-физического 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. Своевременное выявление описанных случаев также возможно с помощью постоянного мониторинга клавиатурного почерка.

- Определение авторства документов, программного кода и т.д. Эта задача актуальна для систем, позволяющих проводить экзамены и тесты онлайн. В случае постоянного подтверждения личности с помощью клавиатурного почерка у кандидата появляется возможность проходить тесты из дома.

В работе [1] указывается, что при печати на клавиатуре у человека задействовано до 140 мышц. Исходя из предположения, что наиболее существенное влияние оказывают 20% от общего числа мышц, получим 28 таких мышц, и соответственно 28-мерную задачу управления. Задача большой размерности позволяет говорить об уникальности клавиатурного почерка пользователей.

1.4 Актуальность исследования

Актуальность и востребованность исследования подтверждается количеством публикаций, которое увеличивается с каждым годом. В работах [4,8] приведен график зависимости количества публикаций от года. На оси

абсцисс изображен год, на оси ординат – количество публикаций. График представлен на рисунке 8.



Рисунок 8 – Количество публикаций по теме исследования

2 Проектирование

2.1 Функциональные возможности системы

Разработанная система может быть эффективно использована для распознавания пользователей. Функциональные возможности включают в себя:

- Возможность получения данных о клавиатурном почерке пользователя во время его работы за компьютером: ВУК, паузы, наложения.
- Возможность регистрации пользователя в системе путем ввода логина и набора 1000 символов в режиме обучения программы.
- Возможность добавления образцов КП к первичному образцу пользователя.
- Возможность определить пользователя по его клавиатурному почерку на основании имеющихся в БД образцов.
- Возможность вести журналирование и записывать в лог-файл когда какой пользователь работал за компьютером.
- Возможность вести мониторинг клавиатурного почерка в скрытом режиме.

Созданное программное приложение предназначено для непрерывного сбора информации о клавиатурных нажатиях на русском/английском языках и идентификации санкционированного или несанкционированного пользователя. Основные этапы работы приложения приведены на рисунке 9.

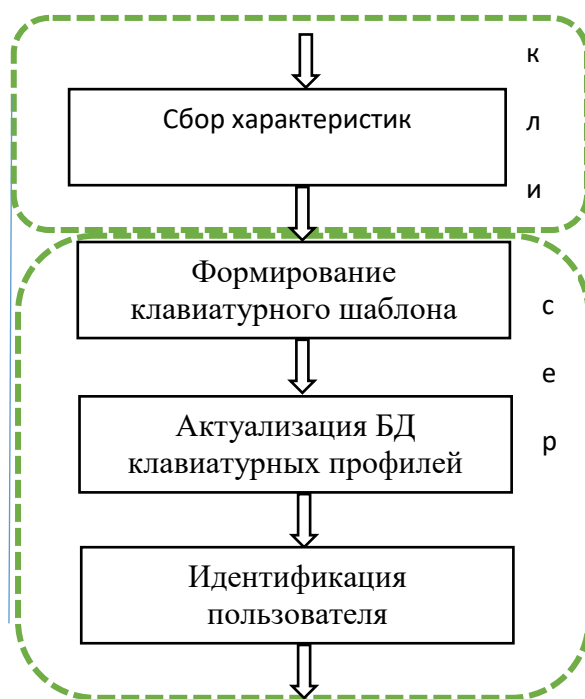


Рисунок 9 – Этапы работы приложения

2.2 Алгоритм работы программы

Алгоритм обучения применяется, когда новый пользователь «регистрируется» в системе, то есть вводит свой логин и оставляет образец своего клавиатурного почерка. Все зарегистрированные системой нажатия клавиш анализируются. Для уменьшения статистической погрешности предлагается считывать сведения о 1000 нажатиях. Для каждой клавиши подсчитывается время удержания, измеряемое в миллисекундах:

$$\text{ВУК} = t_{up} - t_{down} \quad (4)$$

Для каждой буквы русского и английского алфавита, нажатой пользователем, вычисляется среднее время удержания:

$$\overline{\text{ВУК}} = \frac{\sum_{i=1}^n \text{ВУК}_i}{n} \quad (5)$$

где n – общее количество нажатий на клавишу i , произведенных пользователем.

Среднее ВУК рассчитывается отдельно для каждой клавиши с учетом наличия или отсутствия наложений.

Далее происходит расчет пауз между нажатиями:

$$pause = t_{down} - t_{up} \quad (6)$$

Если значение паузы составляет более 1000 мс, это значение отфильтровывается как не подходящее для анализа. Дело в том, что длительная пауза между нажатиями свидетельствует уже не о динамике печати, а об отвлечении пользователя от работы с клавиатурой. Возможно, человек задумался или отвлекся, отошел от компьютера и т.д.

После этого рассчитывается среднее время пауз для образца клавиатурного почерка:

$$\overline{pause} = \frac{\sum_{i=1}^n pause_i}{n} \quad (7)$$

где n – общее количество пауз между нажатиями, за исключением отфильтрованных.

Количество ошибок ввода принимаем равным количеству нажатий на клавиши Delete и Backspace:

$$errors = delete + backspace \quad (8)$$

Пусть P – мера непохожести;

A – значение ВУК-1 в мс;

B – значение ВУК-2 в мс;

N – количество элементов в анализируемой выборке;

S – матрица ковариации.

Эвклидова мера вычисляется по формуле:

$$P = \sqrt{\sum_{i=1}^N (A_i - B_i)^2} \quad (9)$$

Манхэттенское расстояние вычисляется следующим образом:

$$P = \sum_{i=1}^N |A_i - B_i| \quad (10)$$

Расстояние Махаланобиса:

$$P = \sqrt{(\vec{A} - \vec{B})^T S^{-1} (\vec{A} - \vec{B})} \quad (11)$$

Полученные данные записываются в базу данных в качестве образца клавиатурного почерка пользователя. Всего предлагается хранить до 10 образцов почерка для каждого пользователя. Это позволит предусмотреть возможность изменения КП пользователя в зависимости от времени суток и используемой клавиатуры.

Графически алгоритм обучения системы представлен на рисунке 10.

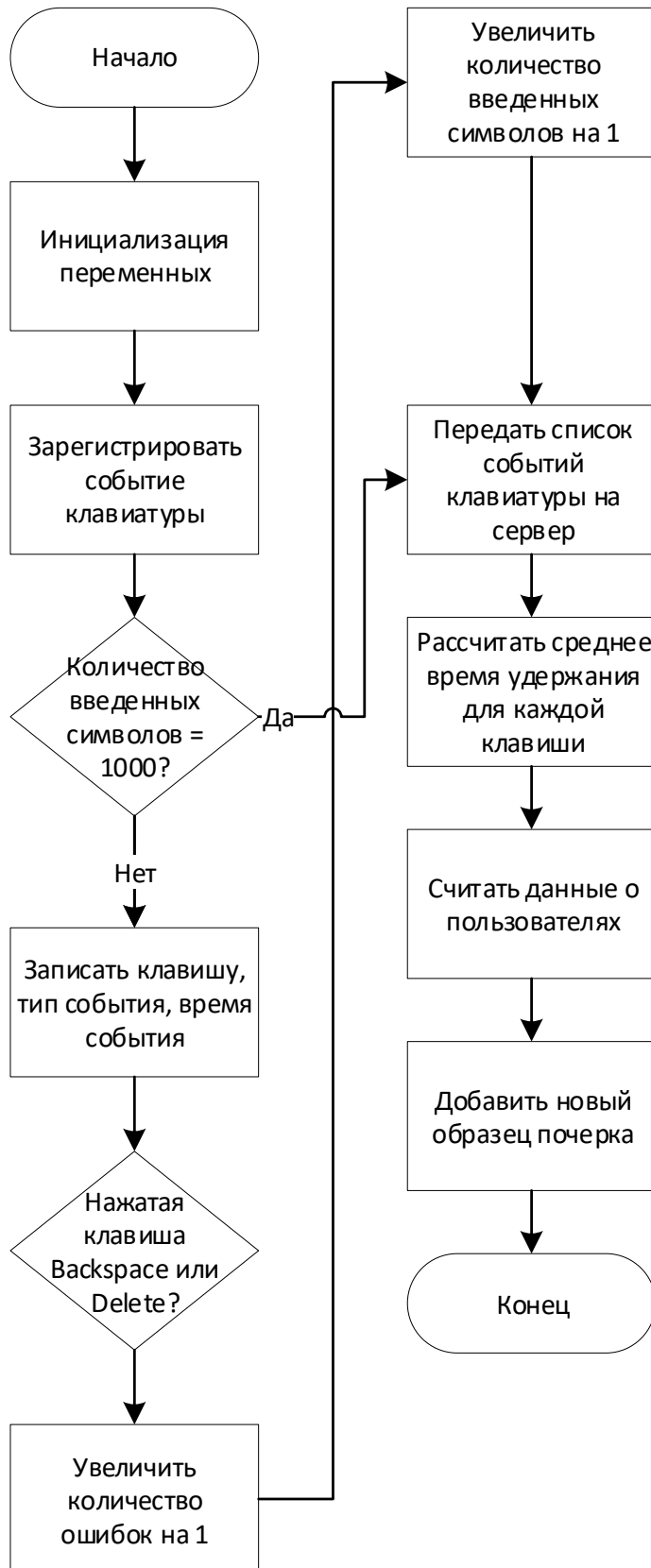


Рисунок 10 – Алгоритм обучения

Также возможно более подробно остановиться на реализации алгоритма подсчета времени удержания клавиш. Он представлен на рисунке 11.

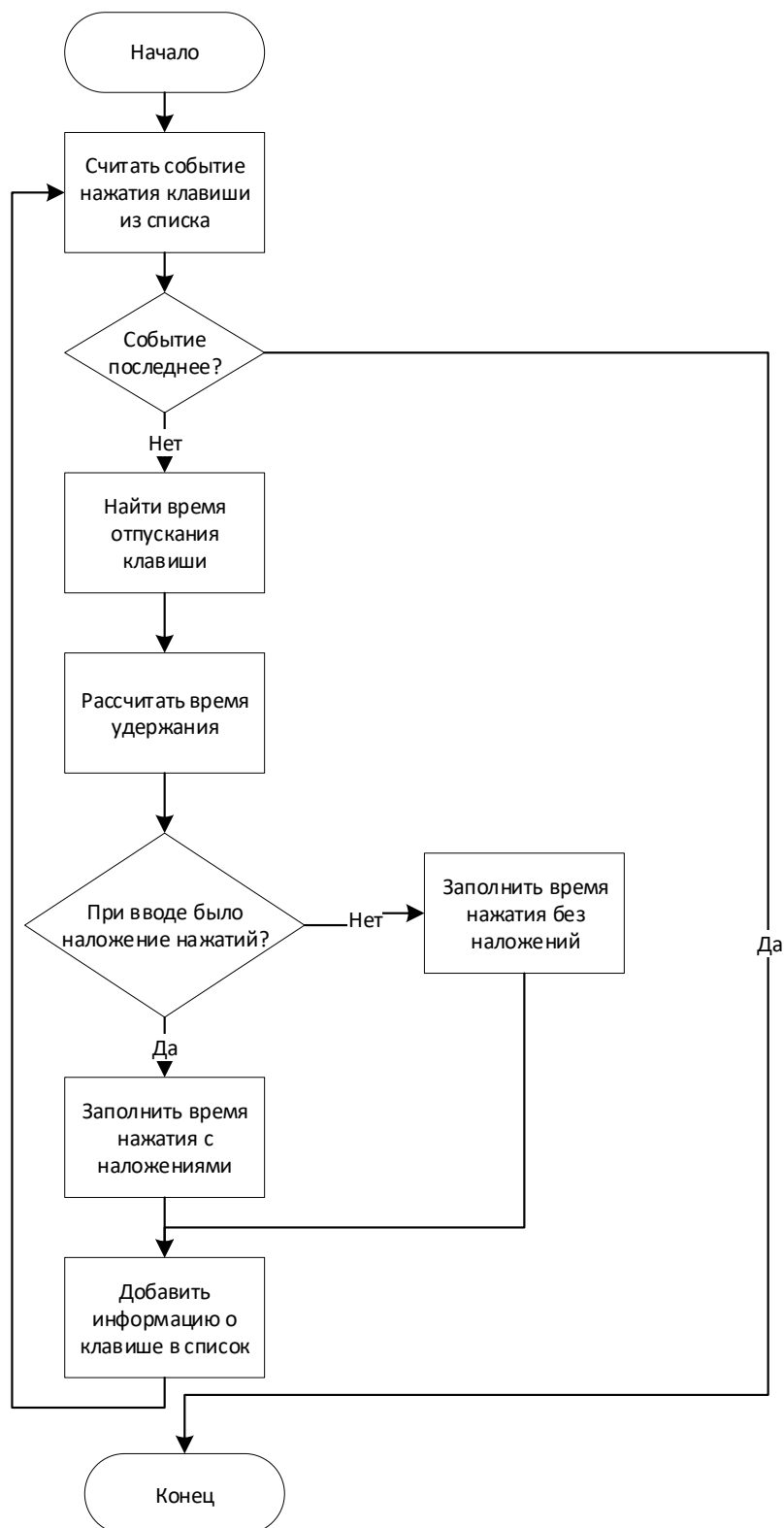


Рисунок 11 – алгоритм подсчета времени удержания клавиш

В алгоритме распознавания считываются сведения о 200 нажатиях на клавиатуру и посылаются на сервер для анализа. Для полученного образца клавиатурного почерка рассчитываются ВУК, длительность пауз, количество ошибок ввода, определяется наличие наложений. После этого полученный образец по очереди сравнивается с каждым образцом КП, имеющимся в базе. Расчет происходит по следующей формуле.

Алгоритм распознавания:

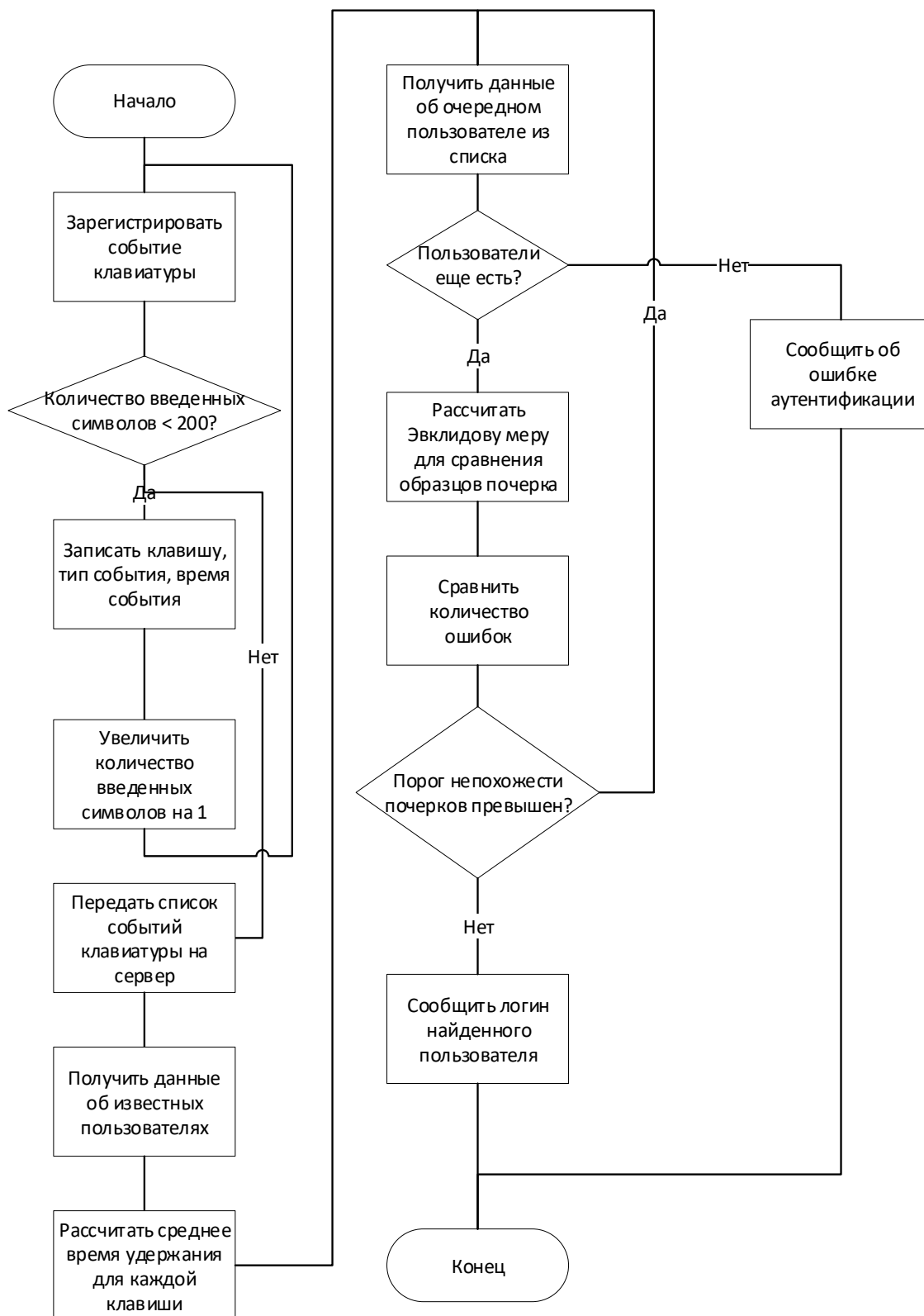


Рисунок 12 – Алгоритм распознавания

Клавиатурный почерк в значительной мере зависит от частоты появления букв алфавита в тексте. Для повышения эффективности работы алгоритма было предложено использовать весовые коэффициенты частоты

букв алфавита при определении расстояния между образцами клавиатурного почерка. В соответствии с рекомендациями ЦКРЯ буквы русского алфавита разделяют на три группы по частоте появления в тексте [24]. В первую группу входят наиболее часто встречающиеся буквы русского алфавита: «О», «Е», «А», «И», «Н», «Т», «С», «Р», «В», «Л», «К». Частота для первой группы варьируется от 3,486% до 10,983%. Во вторую группу входят буквы «М», «Д», «П», «У», «Я», «Ы», «Ь», «Г», «З» и «Б». Их частота встречаемости составляет 1,592% - 3,203%. Третья группа содержит наиболее редко встречающиеся буквы: «Ч», «Й», «Х», «Ж», «Ш», «Ю», «Ц», «Щ», «Э», «Ф», «Ъ» и «Ё». Для них частота встречаемости от 0,013% до 1,45%. При подсчете расстояния между образцами клавиатурного почерка использовались весовые коэффициенты: 0.5 для букв первой группы, 0.3 для букв второй группы, 0.2 для букв третьей группы.

При определении непохожести образцов почерка используются следующие условия:

$$1. \textit{difference} < \textit{Threshold} * \textit{pressingSum}$$

$$|\textit{mistakes} - \textit{averageMistakes}| <$$

$$\textit{Threshold} * \textit{averageMistakes}$$

2.

$$3. |\textit{pauses} - \textit{averagePauses}| < \textit{Threshold} * \textit{averagePauses}$$

где *difference* – Эвклидово расстояние между последовательностями ВУК (для двух сравниваемых образцов почерка) с учетом весовых коэффициентов;

Threshold – пороговое значение, составляет 10%;

pressingSum – сумма ВУК для образца почерка 2;

mistakes – количество ошибок ввода в образце почерка 1;

averageMistake – количество ошибок ввода в образце почерка 2;

pauses – средняя длительность пауз в образце почерка 1;

averagePauses - средняя длительность пауз в образце почерка 2.

При одновременном выполнении всех трех условий образцы почерка считаются совпадающими, то есть принадлежащими одному и тому же пользователю.

2.3 Выбор средств разработки

Для разработки приложения был выбран язык C#. C# – объектно-ориентированный язык программирования от Microsoft, поддерживающий также компонентно-ориентированное программирование. Его синтаксис схож с другими C-подобными языками, например, C++ и Java. В отличие от C++, в C# невозможно множественное наследование классов. C# имеет статическую типизацию, поддерживает такие возможности как [25]:

- Полиморфизм;
- Перегрузку операторов (в том числе операторов явного и неявного приведения типа);
- Делегаты;
- Атрибуты;
- События;
- Свойства;
- Обобщённые типы и методы;
- Анонимные функции с поддержкой замыканий;
- LINQ;
- Исключения;
- Комментарии в формате XML;
- Автоматическая сборка мусора.

Серверная часть приложения реализована в виде консольного приложения. Консольное приложение имеет текстовый интерфейс, то есть

использует при вводе-выводе исключительно буквенно-цифровые символы [26]. Это позволяет обеспечить высокую скорость вывода данных вместе с низкой требовательностью к программным ресурсам. Консольные программы в основном используются, когда нет необходимости в графическом интерфейсе пользователя. Кроме того, они могут отвечать за выполнение рутинных действий пользователя. Серверный компонент не предусматривает получение информации от пользователя через консоль, но отображает ход список подключений через сокеты.

Клиентская часть приложения – это приложение Windows Forms. Windows Forms – технология от Microsoft, которая является частью .NET Framework и отвечает за графический интерфейс пользователя [27]. Windows Forms реализуется в рамках пространства имён System.Windows.Forms. Данная технология упрощает доступ к элементам интерфейса MS Windows путем создания обёртки существующего Win32 API в управляемом коде. При этом классы, реализующие API для Windows Forms, не зависят от языка разработки, что позволяет писать код на C#, C++, VB.Net, J# и других языках. Приложения Windows Forms являются событийно-ориентированными, то есть ожидают действий от пользователя для начала работы. Пользовательский интерфейс при этом включает кнопки, таблицы, списки, переключатели, поля ввода, поля для рисования, изображения, графики. Также возможно отображать сообщения для пользователя с помощью MessageBox.

Для передачи данных между клиентским и серверным компонентом используются TCP сокеты. Сокеты – это название программного интерфейса, обеспечивающего обмен данными между процессами [28]. Сокет — абстрактный объект, представляющий конечную точку соединения. Серверный сокет к определенному TCP порту и слушает его, находясь в режиме ожидания. При этом сохраняется возможность проверить наличие соединений на данный момент или установить тайм-аут для операции.

Серверный сокет обладает уникальным адресом, что позволяет клиенту явно подключиться, установить соединение и передавать данные.

Данные между сервером и клиентом передаются в унифицированном формате JSON. Данные включают сведения о том, когда какая клавиша была нажата или отпущена, сколько было ошибок ввода, а также логин пользователя, время начала и окончания работы.

2.4 Архитектура системы

Для обеспечения возможности одновременной работы программы на нескольких устройствах с общей базой данных, было принято решение реализовать клиент-серверную архитектуру. Данный тип архитектуры обладает следующими достоинствами:

- Снижение требований к клиентским компьютерам, так как основная часть вычислений выполняется на сервере;
- Возможность централизованного хранения данных на сервере обеспечивает лучшую защищенность.

UML диаграмма развертывания иллюстрирует архитектуру разработанного приложения.

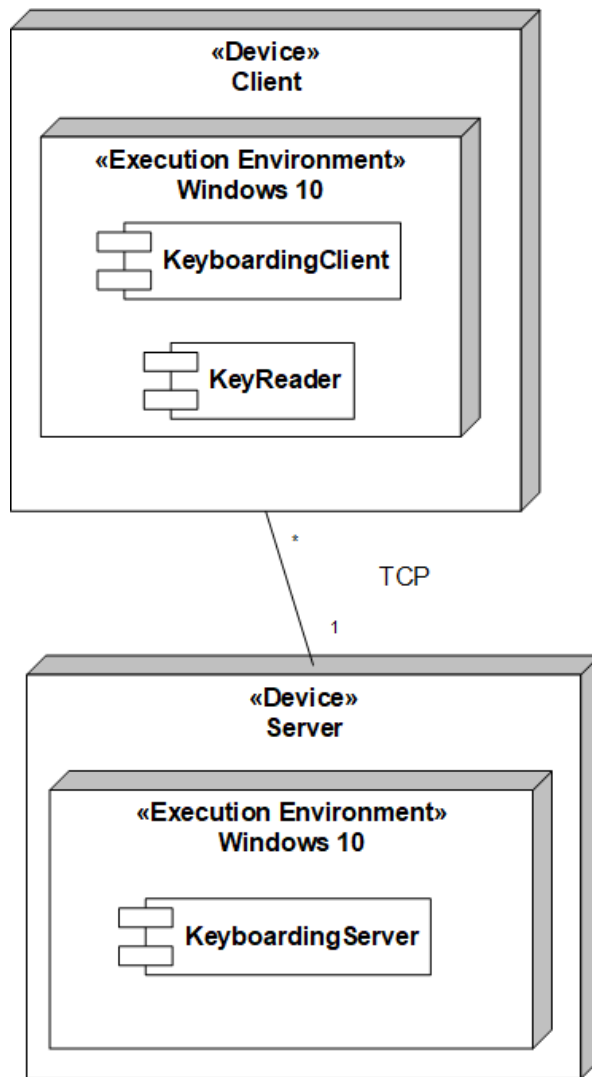


Рисунок 13 – Диаграмма развертывания

Более детально архитектуру можно представить на диаграмме классов. Диаграмма классов для серверной части показана на рисунке 14, а диаграмма классов для клиентской части – на рисунке 15.

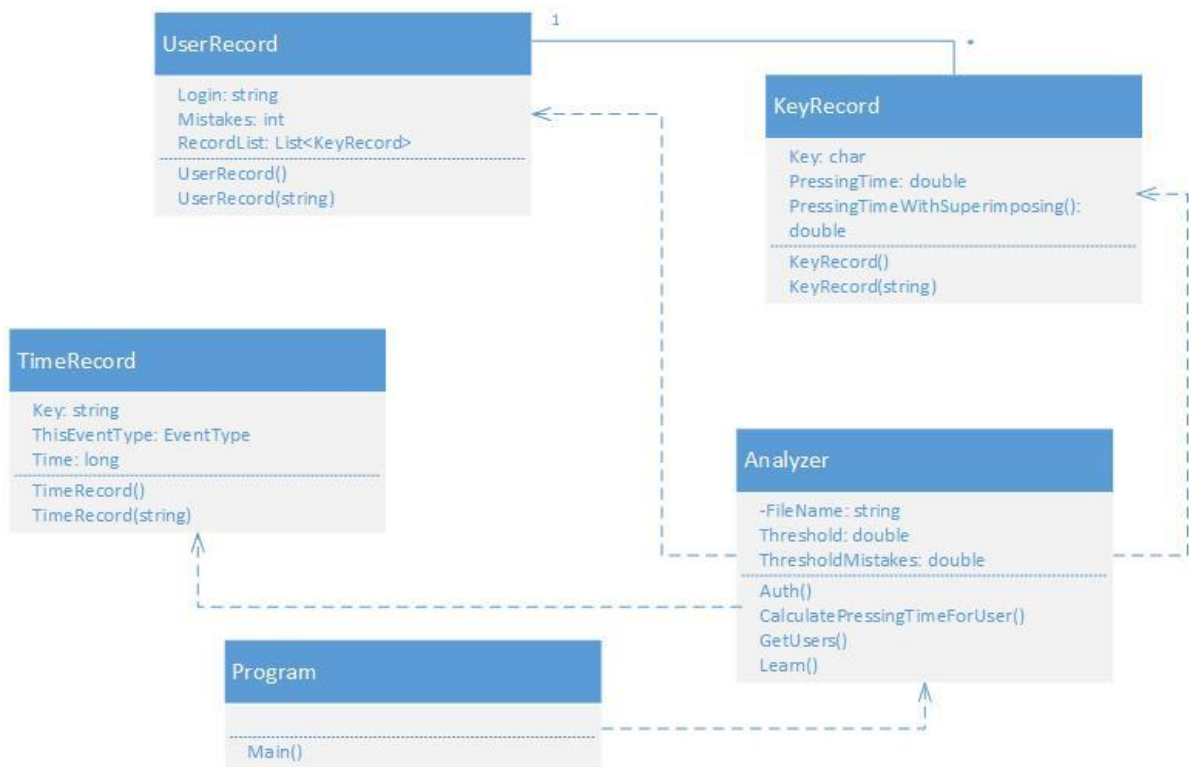


Рисунок 14 – Диаграмма классов серверного приложения

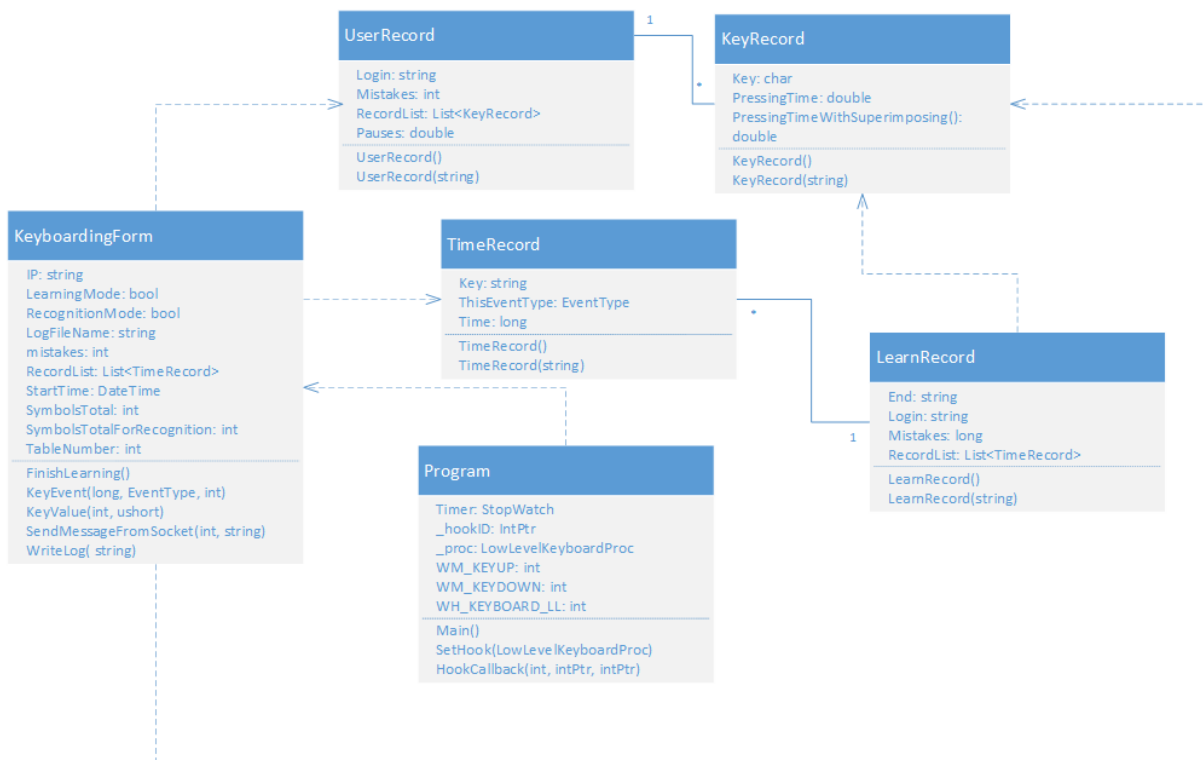


Рисунок 15 – Диаграмма классов клиентского приложения

3 Реализация и тестирование

3.1 Описание интерфейса

При первом запуске пользователь увидит экран, представленный на рисунке 16.

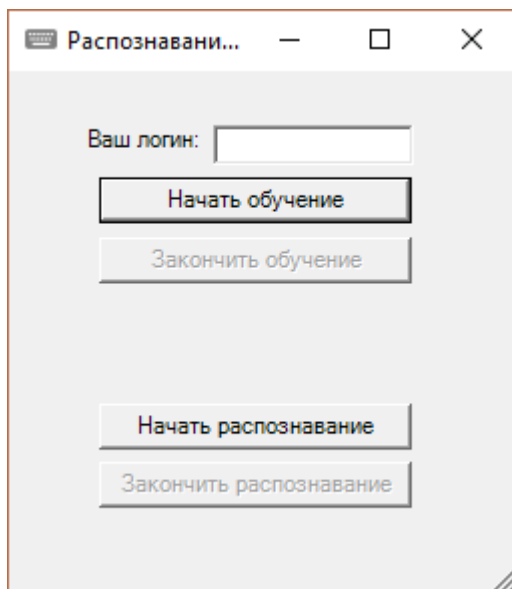


Рисунок 16 – Приветственный экран

В поле «Ваш логин» необходимо ввести логин, по которому система в дальнейшем будет узнавать пользователя. Если нажать кнопку «Начать обучение», не выбрав логин, система выдаст ошибку, показанную на рисунке 17.

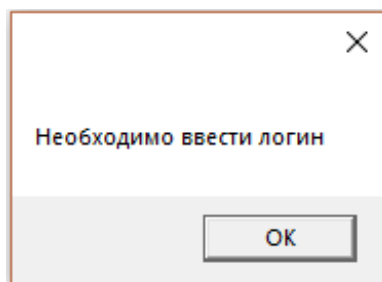


Рисунок 17 – Сообщение о необходимости ввести логин

После ввода логина, необходимо нажать кнопку «Начать обучение». После этого появляется информация о количестве введенных символов и общем количестве символов, необходимых для завершения процесса обучения. Информация о количестве введенных символов обновляется с

каждым нажатием. Стоит отметить, что во время обучения пользователь не имеет возможности изменить свой логин – данное поле ввода не активно.

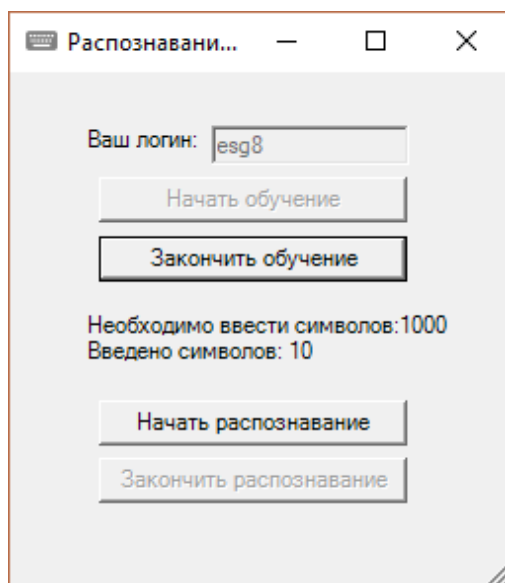


Рисунок 18 – Процесс обучения

После того, как количество введенных символов достигает 1000, пользователю показывается сообщение следующего вида.

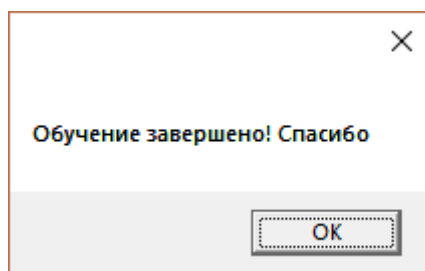


Рисунок 19 – Сообщение об окончании обучения

В случае если обучение было досрочно прервано, выдается другое сообщение, показанное на рисунке.

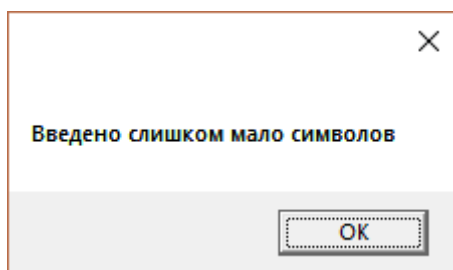


Рисунок 20 – Сообщение о досрочном окончании обучения

Другой режим работы приложения – распознавание пользователя – можно запустить вручную из интерфейса программы, нажав кнопку «Начать

распознавание». После этого станет доступна кнопка «Закончить распознавание». До того, как пользователь нажмет на кнопку «Закончить распознавание», программа будет собирать данные о нажатиях и отправлять на сервер порциями по 200 символов. На рисунке 21 представлен интерфейс работы в режиме распознавания.

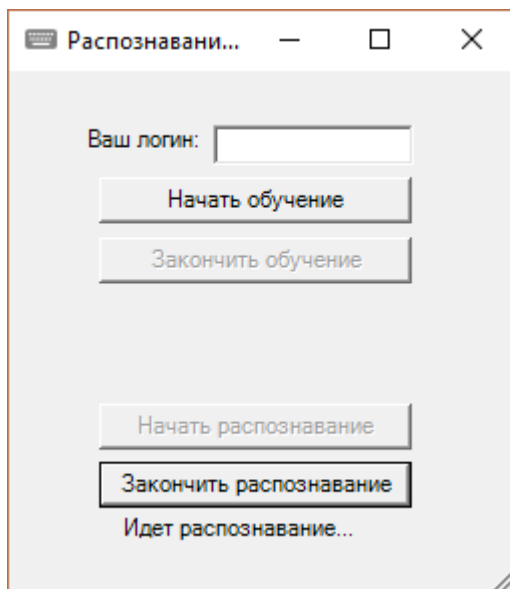


Рисунок 21 – Процесс распознавания

Также программа позволяет производить скрытый мониторинг. В этом случае пользователю не нужно нажимать на кнопку «Начать распознавание» или «Закончить распознавание». Приложение постоянно запущено в трее и непрерывно собирает сведения о клавиатурных нажатиях. Для этого его следует добавить в автозапуск, например, администратору корпоративной сети.

3.2 Сбор тестовых данных

Первый этап сбора тестовых данных включал использование созданной программы в режиме обучения. При работе с программой сведения о клавиатурных почерках сохраняются базу данных. При этом возможна одновременная работа нескольких пользователей.

Для каждого почерка анализируются и сохраняются следующие характеристики:

- Среднее время удержание каждой клавиши;

- Среднее время удержание каждой клавиши при наличии наложений;

- Среднее время паузы между нажатиями;

- Количество ошибок ввода.

Также сохраняется логин пользователя для возможности его дальнейшей идентификации.

На втором этапе, в соответствии с лицензионным соглашением, была получена база данных клавиатурных почерков из Мадрида [5, 29]. Эта база данных включает сведения о клавиатурных почерках 300 пользователей, 24 образца для каждого пользователя.

Данные представлены в виде текстовых файлов с названиями вида TXXX_YY.txt, где XXX – номер пользователя, YY – номер образца почерка данного пользователя. Файл включает следующую информацию:

23

0x2aD 0

0x24D 265

0x24U 109

0x2aU 0

...

Здесь 23 – количество клавиш,

0x2a – код клавиши из таблицы соответствия,

D – тип события (D – нажатие, U – отпускание),

0 – время между двумя последующими событиями, мс. Первое время всегда 0. В примере время между нажатием клавиши 0x2a и нажатием 0x24 составило 265 мс.

Был написан скрипт на языке C#, конвертирующий данные в формат, пригодный для работы созданной программы.

клавиатурных почерках пользователей Томска. Данные приведены в единый формат и сохранены в общей базе данных.

3.3 Тестирование программы

Тестирование разработанной методики проводилось в несколько этапов.

На первом этапе были исследованы образцы клавиатурного почерка 15 пользователей, собранные разработанным программным приложением. В основном образцы принадлежат студентам ТПУ. Для каждого почерка было подсчитано ВУК без наложений, ВУК с наложениями, паузы, количество ошибок. На рисунке 22 представлено изменение ВУК без наложений в зависимости от нажатой клавиши для пользователя eak23.



Рисунок 22 – ВУК без наложений для пользователя eak23

На рисунке 23 представлено изменение ВУК с наложениями в зависимости от нажатой клавиши для того же пользователя.

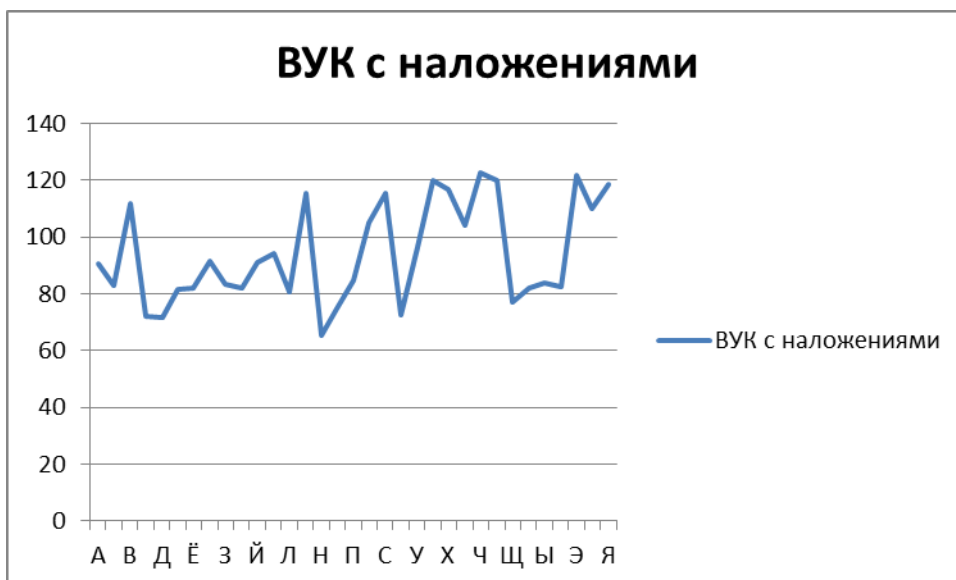


Рисунок 23 – ВУК с наложениями для пользователя eak23

Сравнение ВУК с наложениями и без наложений для одной и той же клавиши показано на рисунке 24.

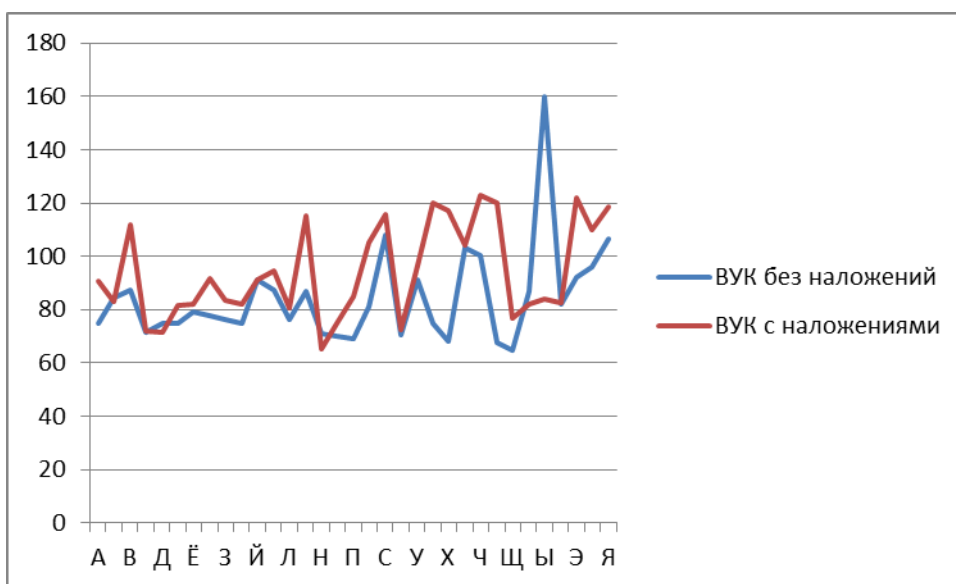


Рисунок 24 – ВУК с наложениями и без наложений для пользователя eak23

Средняя длительность пауз для пяти различных пользователей представлена на рисунке 25.

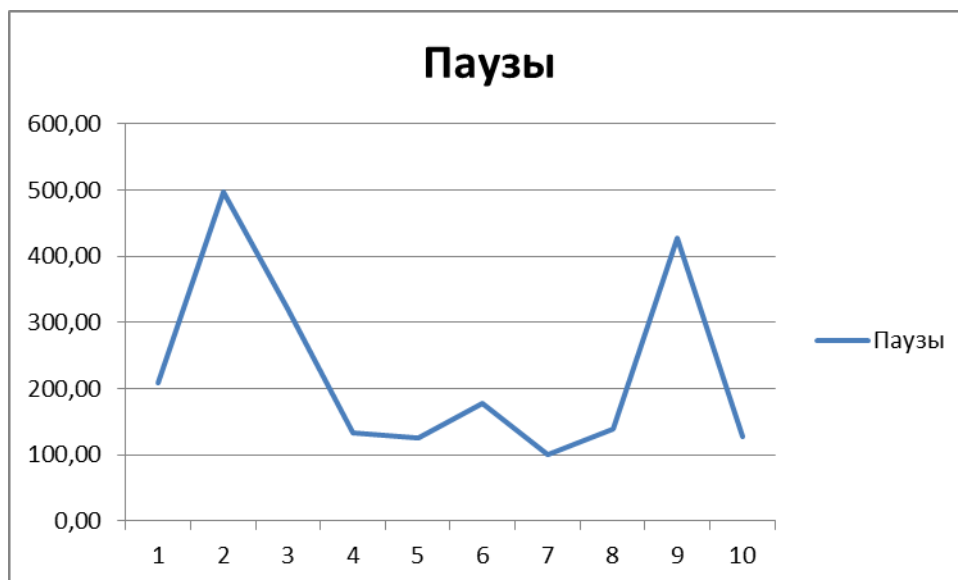


Рисунок 25 – Средняя длительность пауз

На графике 26 показано среднее ВУК для нескольких образцов почерка одного и того же пользователя.

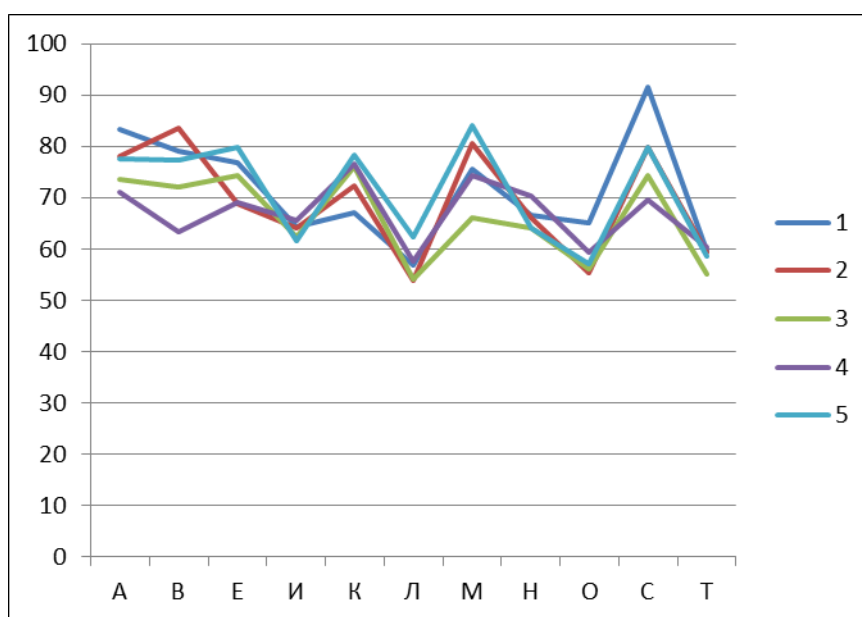


Рисунок 26 – Среднее ВУК для нескольких образцов почерка одного пользователя

На графике 27 представлено среднее ВУК для нескольких пользователей.

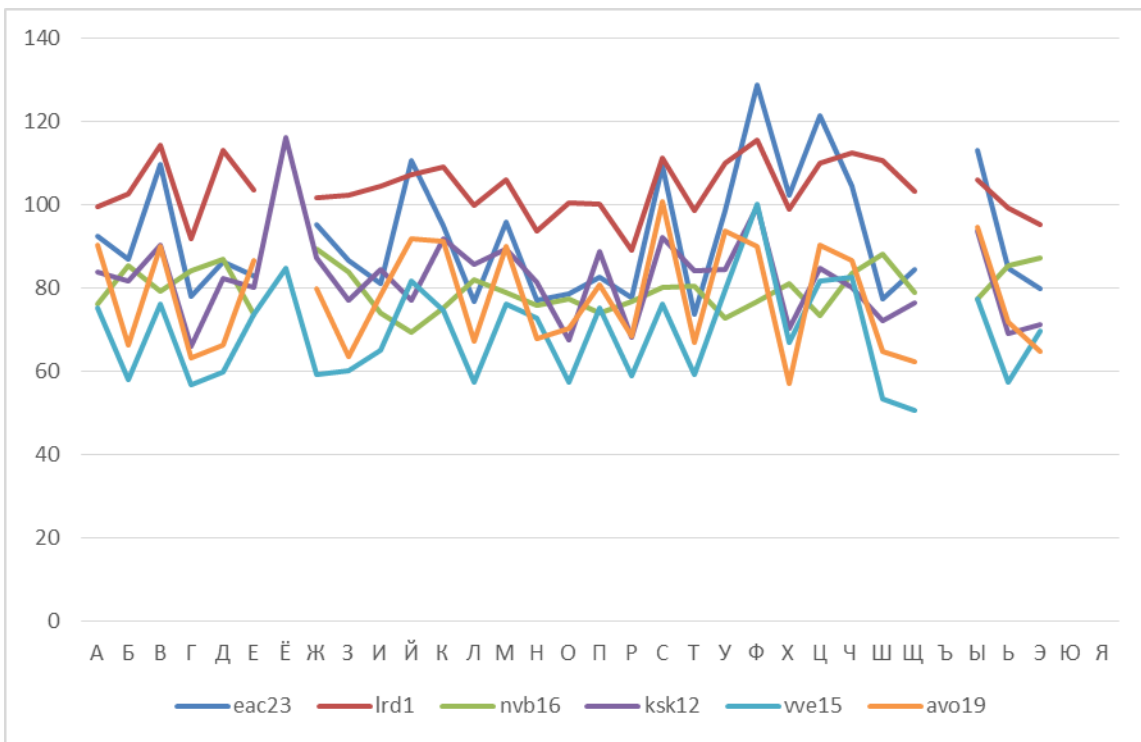


Рисунок 27 – Среднее ВУК для нескольких пользователей

На графике 28 представлено ВУК для одного и того же пользователя в обычном состоянии и в состоянии алкогольного опьянения.

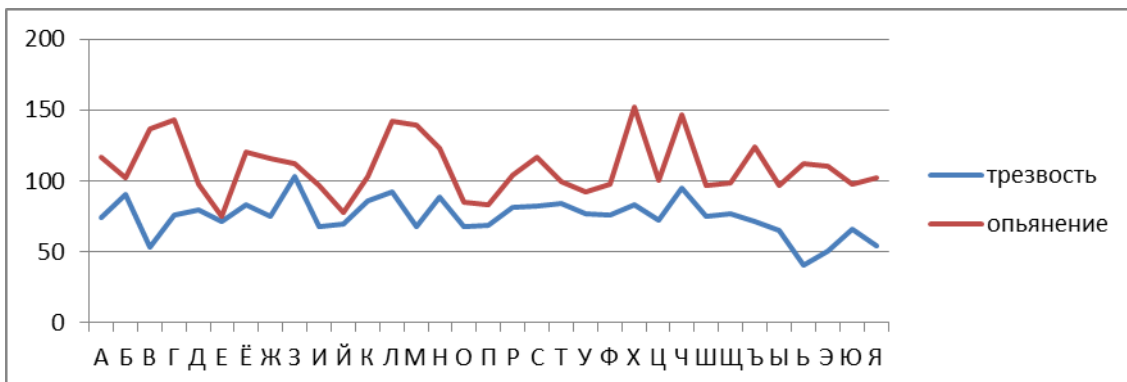


Рисунок 28 – ВУК для пользователя в обычном состоянии и состоянии опьянения

На графике 29 показано ВУК для клавиши «А» для нескольких пользователей.

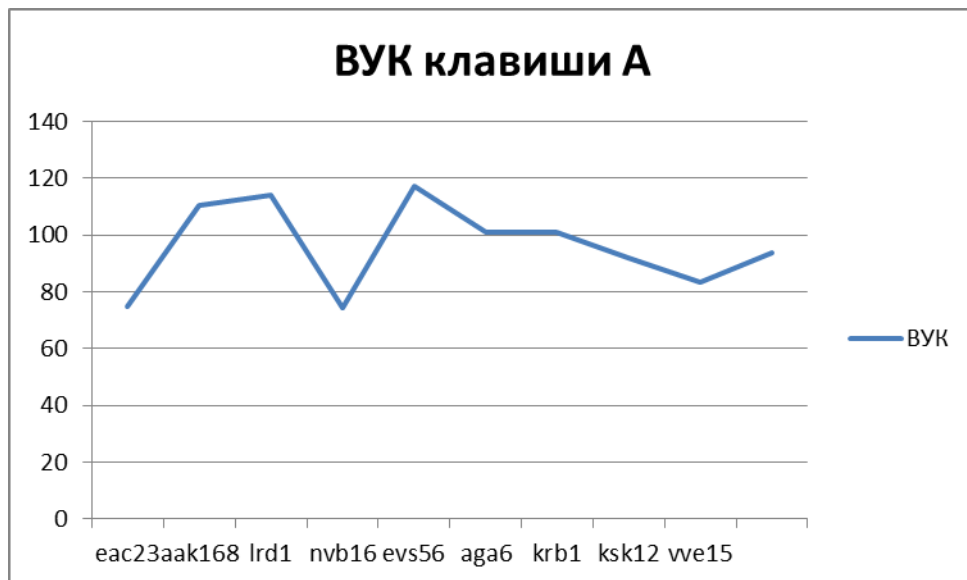


Рисунок 29 – ВУК для клавиши «А»

Таким образом, ВУК лежит в пределах от 74,63мс до 117,47мс. Среднее значение равно 95,51мс.

Полученные результаты позволяют сделать вывод, что клавиатурный почерк разных людей различается достаточно сильно, чтобы была возможность проводить аутентификацию.

На втором этапе тестирования были проведены эксперименты с целью выбора наилучшего порога чувствительности. Эксперимент был проведен над 25 образцами клавиатурного почерка, принадлежащих пяти пользователям из базы данных Мадрида. Для каждого образца было подсчитано среднее время удержания для каждой клавиши. Затем образцы почерков поочередно сравнивались друг с другом. Была вычислена Евклидова мера для каждой пары образцов. Для установления значений порога сходства в ходе эксперимента его значение варьировалось в диапазоне от 0.01 до 0.5 с шагом 0.01 и подсчитано количество ошибок первого и второго рода, а также их сумма. Результат представлен на рисунке 30.

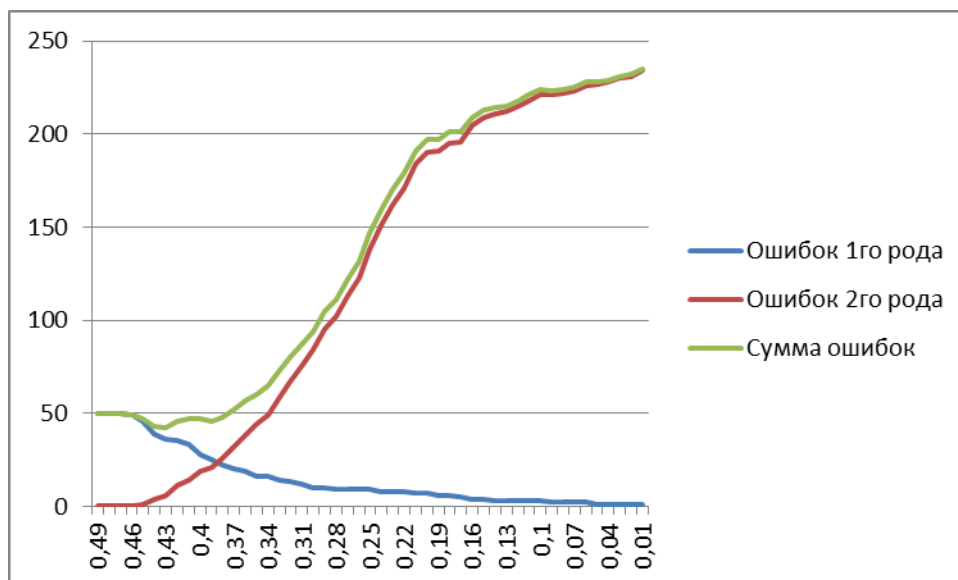


Рисунок 30 – Количество ошибок 1го и 2го рода для английских букв

Аналогичный эксперимент был проведен для образцов КП на русском языке. Результат показан на рисунке 31.

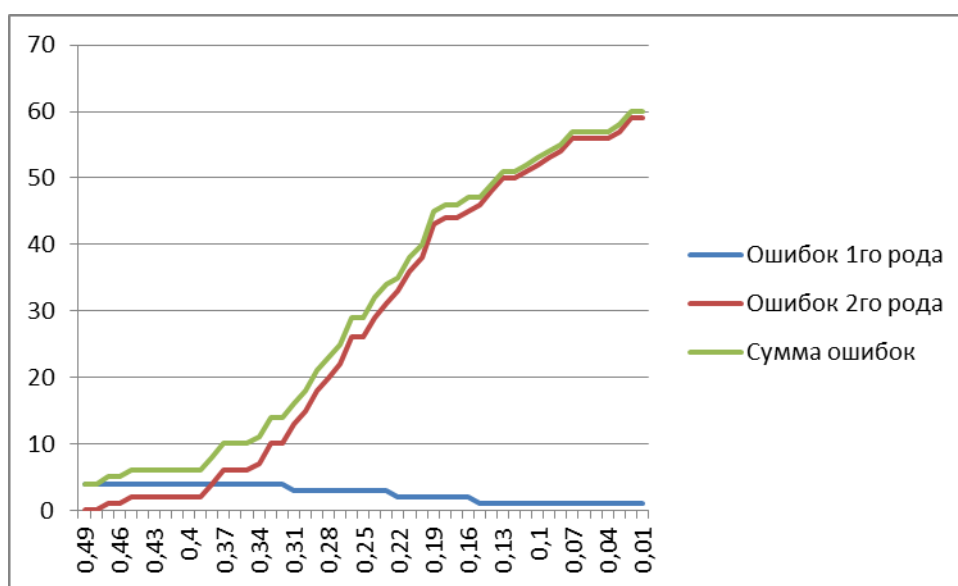


Рисунок 31 – Количество ошибок 1го и 2го рода для русских букв

В таблице 2 проведено сравнение эффективности реализованного алгоритма с другими известными алгоритмами.

Автор	Год	Метод	FAR (%)	FRR (%)
R.Joyce, G.Gupta	1990	Мат.ожидание и вариация	0,25	16,36
S.Bleha, С. Slivinsky	1990	Байесовский классификатор	2,8	8,1

M. S. Obaidat	1993	Персептрон	8	9
J. A. Robinson, V. M. Liang	1998	Индуктивный обучающийся классификатор	9	10
E. Yu, S. Cho	2003	Ассоциативный многослойный персептрон	0	0,814
K. Revett, S. T. deMagalhaes	2007	Алгоритм последовательного приближения	0,15	0,2
Горохова Е.С.	2018	Векторный динамический критерий	1,1	1,9

На третьем этапе тестирования проводились эксперименты с пользователями. 15 пользователей, чьи образцы клавиатурного пользователя были получены ранее, а также 5 неизвестных системе пользователей («шпионов») по очереди работали за компьютером с включенной системой распознавания клавиатурного почерка. При этом велась запись, когда какой пользователь работал за компьютером. После эксперимента логи, созданные программой были сравнены с записями. Для каждого периода времени, таким образом, было возможно определить, какой пользователь действительно работал за компьютером, и какого пользователя система опознала по почерку. В результате было подсчитано количество ошибок 1го и 2го рода. Результаты приведены в таблице 3.

Таблица 3 – Результаты тестирования при участии 20 пользователей

Время	Реальный пользователь	Пользователь, определенный системой
12.15-12.25	esg8	esg8
12.25-12.30	Неизвестный пользователь	Неизвестный пользователь
12.30-12.45	nata	nata
12.45-12.50	kikaswan	kikaswan
12.50-12.55	kanareyka	esg8
12.55-13.05	pavel	pavel
13.05-13.10	anastasia	anastasia

13.10-13.15	Неизвестный пользователь	Неизвестный пользователь
13.15-13.20	alex	alex
13.20-13.25	arsen	arsen
13.25-13.30	dima	dima
13.30-13.40	dana	dana
13.40-13.45	andrey	Неизвестный пользователь
13.45-13.50	ayamila	ayamila
13.50-13.55	tan5	tan5
13.55-14.00	kate	kate
14.00-14.10	Неизвестный пользователь	Неизвестный пользователь
14.10-14.15	Неизвестный пользователь	Неизвестный пользователь
14.15-14.20	sasha	sasha
14.20-14.25	Неизвестный пользователь	Неизвестный пользователь

Таким образом, пользователь был верно определен в 90% случаев.
Была обнаружена 1 ошибка 1го рода и 1 ошибка второго рода.

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) Linux OS;
 - 3) Mac 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 с позиции ресурсоэффективности и ресурсосбережения

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

1. Программное приложение «Клавиатурный почерк 1.0»;
2. Система контроля сотрудников «Стахановец», включающая модуль для распознавания клавиатурного почерка.

Сравнение технических и экономических характеристик этих продуктов представлено в таблице 4. «Клавиатурный почерк 1.0» обозначен К1, а «Стахановец» - К2.

Таблица 4 – Оценочная карта сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Удобство в эксплуатации	0,18	4	1	5	0,72	0,18	0,9
2. Функциональные возможности	0,15	3	1	5	0,45	0,15	0,75
3. Потребность в ресурсах памяти	0,07	5	5	2	0,35	0,35	0,14
4. Быстродействие	0,1	5	5	4	0,5	0,5	0,4
5. Возможность подключения в сеть ЭВМ	0,11	4	1	5	0,44	0,11	0,55
Экономические критерии оценки эффективности							
1. Возможность доработки	0,06	5	3	4	0,3	0,18	0,24
2. Уровень проникновения на рынок	0,1	1	2	4	0,1	0,2	0,4
3. Цена	0,2	4	5	1	0,8	1	0,2
4. Предполагаемый срок эксплуатации	0,01	5	3	4	0,05	0,03	0,04
5. Послепродажное обслуживание	0,02	3	1	5	0,06	0,02	0,1
Итого	1				3,77	2,72	3,72

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

- Цена;

- Потребность в ресурсах памяти;
- Быстродействие;
- Предполагаемый срок эксплуатации;
- Возможность подключения в сеть ЭВМ;
- Удобство в эксплуатации.

Недостатками системы являются:

- Уровень проникновения на рынок;
- Функциональные возможности;
- Послепродажное обслуживание.

В результате анализа установлено, что конкурентоспособность системы выше, чем у К1 и К2, соответственно, целесообразно проводить разработку по данной тематике.

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 были описаны сильные и слабые стороны проекта, возможности и угрозы реализации. Они представлены в таблице А.1 в приложении А.

Второй этап SWOT-анализа включает выявление соответствий между сильными и слабыми сторонами проекта и окружающей средой. Интерактивные матрицы соответствия представлены в таблицах 5-8.

Таблица 5 – Интерактивная матрица соответствия сильных сторон и возможностей

		Сильные стороны проекта					
		C1	C2	C3	C4	C5	C6
Возможности проекта	B1	-	+	+	0	-	+
	B2	-	+	+	+	-	-
	B3	+	-	+	+	+	+
	B4	-	+	-	-	-	-
	B5	0	+	-	+	-	-

Таблица 6 – Интерактивная матрица соответствия сильных сторон и угроз

Сильные стороны проекта							
Угрозы проекта		C1	C2	C3	C4	C5	C6
	У1	-	-	-	-	0	-
	У2	-	-	-	-	-	-
	У3	-	+	-	-	-	-

Таблица 7 – Интерактивная матрица соответствия слабых сторон и возможностей

Возможности проекта		Сл1	Сл2	Сл3
	В1	+	-	+
	В2	-	-	+
	В3	-	-	-
	В4	+	-	+
	В5	-	+	+

Таблица 8 – Интерактивная матрица соответствия слабых сторон и угроз

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

На третьем этапе составляется итоговая матрица SWOT-анализа. Она представлена в таблице А.2 в приложении А.

4.1.4 Оценка готовности проекта к коммерциализации

Оценка готовности выполнена путем ответов на вопросы, представленные в таблице А.3 в приложении А. Каждый ответ дает от 1 до 5 баллов.

Таким образом, готовность научного проекта к коммерциализации выше среднего. Уровень имеющихся знаний у разработчика немного выше, но также находится в категории выше среднего. В дальнейшем необходимо больше внимания уделить вопросам выхода на зарубежный рынок и оценке стоимости интеллектуальной собственности.

4.1.5 Методы коммерциализации результатов научно-технического исследования

Наиболее подходящим методом коммерциализации проекта по разработке системы распознавания клавиатурного почерка является передача ноу-хау. Были выделены следующие причины:

- Разработка системы распознавания клавиатурного почерка началась по предложению коммерческой организации, заинтересованной в применении создаваемой программной системы;

- Разработчик не имеет опыта, чтобы запустить торговлю патентными лицензиями самостоятельно. Также отсутствует клиентская база;

- Передача ноу-хау является быстрым методом коммерциализации.

4.2 Инициация проекта

В данном разделе представлен устав магистерской работы. В таблице 9 показаны заинтересованные стороны и их ожидания.

Таблица 9 – Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Разработчик системы	Получение знаний по специальности, пополнение портфолио, получение материальной выгоды
НИ ТПУ	Увеличение числа научных публикаций, дипломов на научно-практических конференциях
IT-компания	Получение ноу-хау по распознаванию клавиатурного почерка, получение материальной выгоды

В таблице 10 рассмотрены цели, ожидаемые результаты проекта, а также критерии достижения целей.

Таблица 10 – Цели и результаты проекта

Цели проекта:	Разработка методики распознавания пользователю по клавиатурному почерку, а также программной системы на основе этой методики. Коммерциализация проекта путем передачи ноу-хау.
Ожидаемые	Работающая система распознавания пользователя по

результаты проекта:	клавиатурному почерку.
Критерии приемки результата проекта:	Система может использоваться несколькими пользователями одновременно. Точность распознавания составляет не менее 90%.
Требования к результату проекта:	Требование:
	Система распознавания клавиатурного почерка должна работать в локальной сети.
	Система распознавания клавиатурного почерка должна поддерживать одновременную работу нескольких пользователей.
	При распознавании пользователей точность алгоритма должна составлять не менее 90%.

4.2.1 Ограничения и допущения проекта.

При разработке системы распознавания клавиатурного почерка необходимо учитывать несколько ограничений. Они представлены в таблице 11.

Таблица 11 – Ограничения проекта

Фактор	Ограничения/ допущения
3.1. Бюджет проекта	Бюджет ограничен суммой в 50000 рублей
3.1.1. Источник финансирования	Личные средства разработчика
3.2. Сроки проекта:	01.09.2016-30.05.2018
3.2.1. Дата утверждения плана управления проектом	01.10.2016
3.2.2. Дата завершения проекта	30.05.2018
3.3. Прочие ограничения и допущения	Время работы участников проекта не может превышать

4 часа в день

4.3 Планирование управления научно-техническим проектом

4.3.1 План проекта

План проекта представлен в виде диаграммы Ганта. Он приведен в таблице А.4 в приложении А.

4.3.2 Бюджет научного исследования

В данном разделе рассматриваются вопросы планирования бюджета НТИД.

Расходы на сырье и материалы представлены в таблице 12.

Таблица 12 – Сырье, материалы, комплектующие изделия и покупные полуфабрикаты

Наименование	Марка, размер	Кол-во	Цена за единицу, руб.	Сумма, руб.
Бумага для офисной техники	Снегурочка, А4	1	229	229
Ручка шариковая	Attache Style	3	25	75
Всего за материалы				304
Транспортно-заготовительные расходы (3-5%)				9,12
Итого по статье C_m				313,12

Для выполнения работы необходим персональный компьютер. В таблице 13 представлены материальные затраты НТИ с учетом транспортных расходов.

Таблица 13 – Расчет затрат по статье «Спецоборудование для научных работ»

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб.
1.	Компьютер	1	39,5	39,5
2.	Visual Studio 2017 Community	1	-	-

Основная заработная плата представлена в таблице 14.

Таблица 14 – Расчет основной заработной платы

№ п/п	Наименование этапов	Исполнители по категориям	Трудо-емкость, чел.-дн.	Заработная плата, приходящаяся на один чел.-дн., тыс.руб.	Всего заработная плата по тарифу (окладам), тыс. руб.
1	Обзор существующего ПО, обзор существующих методик, разработка методики и алгоритма, реализация изменений, сбор тестовых данных, тестирование продукта, оценка эффективности полученных результатов, составление пояснительной документации	Студент	80	0,273	21,84
2	Обзор существующих методик, разработка методики и алгоритма, оценка эффективности полученных результатов	Руководитель	37	1,371	50,727
Итого: 72,567					

Расчет дополнительной заработной платы представлен в таблице 15.

Таблица 15 – Расчет дополнительной заработной платы

Заработная плата	Руководитель	Студент
Основная зарплата	50,727	21,84

Дополнительная зарплата	6,087	2,621
Итого по статье С _{зп}	8,708	

Накладные расходы включают прочие затраты на выполнение НИИ, в частности, затраты на электроэнергию, печать и ксерокопирование, услуги связи и т.д. Величину накладных расходов примем за 80% от основной и дополнительной заработной платы, что составляет 65200 рублей.

Отчисления на социальные нужды в ПФРФ составляют $0,22 * (72,567 + 8,708) = 17,8805$ тыс.руб.

Отчисления в ФСС составляют $0,029 * (72,567 + 8,708) = 2,3569$ тыс.руб.

Отчисления в ФФМС составляют $0,051 * (72,567 + 8,708) = 4,145$ тыс.руб.

Итого во внебюджетные фонды необходимо отчислить 24,3824 тыс.руб.

Расходы на командировки научного и производственного персонала принимаются равными $0,1 * (72,567 + 8,708) = 8,1275$ тыс.руб.

Для выполнения НИИ требуется постоянный доступ в интернет, поэтому оплата стоимости работ, выполняемых сторонними организациями и предприятиями, рассчитывается как $5 * 0,6 = 3$ тыс.руб.

Группировка затрат по статьям представлена в Приложении А в таблице А.5.

4.3.2 Организационная структура проекта

Возможны три варианта организационной структуры проекта: функциональная, матричная и проектная.

Для реализации НИИ была выбрана матричная структура, поскольку проект технологический сложный, время исполнения ограничено, взаимозависимость проекта от организаций более высокого уровня средняя.

Схема структуры проекта представлена на рисунке 32.

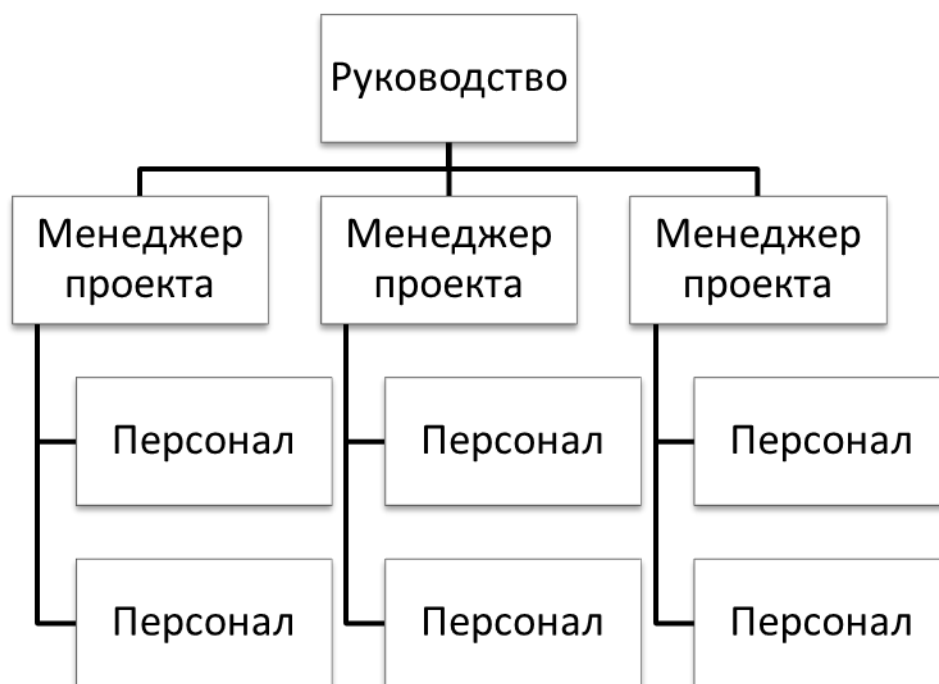


Рисунок 32 – Схема структуры проекта

4.3.3 План управления коммуникациями проекта

План управления коммуникациями отражает требования к коммуникациям со стороны участников проекта. Он представлен в таблице 16.

Таблица 16 – План управления коммуникациями

№ п/п	Какая информация передается	Кто передает информацию	Кому передается информация	Когда передает информацию
1.	Статус проекта	Исполнитель проекта	Руководителю проекта	Ежеквартально (первая декада квартала)
2.	Обмен информацией о текущем состоянии проекта	Исполнитель проекта	Участникам проекта	Еженедельно (вторник)
3.	Документы и информация по проекту	Ответственное лицо по направлению	Руководителю проекта	Не позже сроков графиков и к. точек
4.	О выполнении контрольной точки	Исполнитель проекта	Руководителю проекта	Не позже дня контрольного события по плану управления

4.3.4 Реестр рисков проекта

Риски проекта включают в себя различные неопределенные события, которые могут возникнуть в проекте и вызвать негативные последствия. Риски представлены в таблице 17.

Таблица 17 – Реестр рисков

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска *	Способы смягчения риска	Условия наступления
1	Несоблюдение сроков проекта	Проект не будет закончен вовремя	2	4	Средний	Составление плана реализации и проекта	Неверное планирование времени
2	Недостаток знаний не позволит создать продукт, отвечающий требованиям	Продукт ненадлежащего качества	2	5	Средний	Чтение литературы по специальности	Недостаток знаний и навыков у разработчика
3	Создание продукта, не соответствующего ожиданиям заказчика	Невостребованность системы	1	2	Низкий	Сбор требований и тщательное проектирование системы	Изменение требований к системе

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

4.4.1 Оценка абсолютной эффективности исследования

В таблице А.6 в приложении А представлен расчет чистой текущей стоимости (NPV).

Таким образом, чистая текущая стоимость по проекту в целом составляет 223 тыс.руб., что позволяет судить о его эффективности.

Рассчитаем дисконтированный срок окупаемости.

Таблица 18 – Дисконтированный срок окупаемости.

№	Наименование показателя	Шаг расчета				
		0	1	2	3	4
1.	Дисконтированный чистый денежный поток ($i=0,20$)	-222	162	135	113	94
2.	То же нарастающим итогом	-280	-118	17	130	223
3.	Дисконтированный срок окупаемости	$PP_{диск} = 1 + 118/135 = 1,87$ года				

Рассчитаем внутреннюю ставку доходности IRR в таблице А.7 в приложении А.

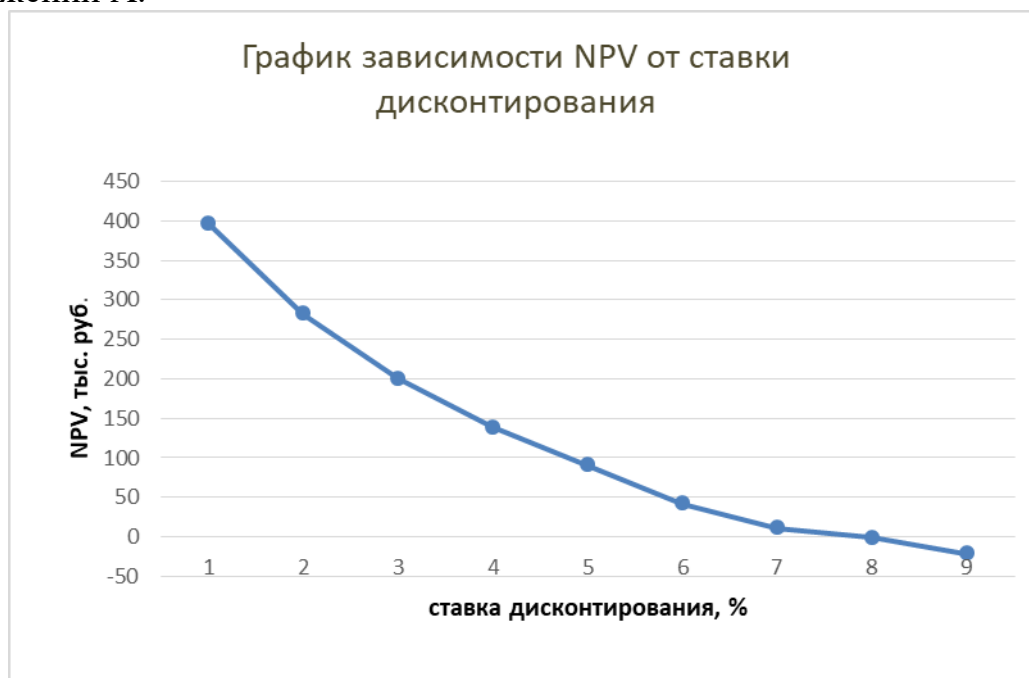


Рисунок 33 – Зависимость NPV от ставки дисконтирования

Из таблицы и графика видно, что по мере роста ставки дисконтирования чистая текущая стоимость уменьшается, становясь отрицательной. Значение ставки, при которой NPV обращается в нуль, носит название «внутренней ставки доходности» или «внутренней нормы прибыли». Из графика получаем IRR составляет 0,75.

Рассчитаем индекс доходности (рентабельности) инвестиций по формуле:

$$PI = \sum_{t=1}^n \frac{ЧПД_t}{(1+i)^t} / I_0, \quad (12)$$

где I_0 – первоначальные инвестиции.

$$PI = \frac{162 + 135 + 113 + 94}{221} = 2,28$$

$PI=2,28>1$, следовательно, проект эффективен при $i=0,2$; $NPV=223$ тыс. руб.

Оценка социальной эффективности представлена в таблице 19.

Таблица 19 – Критерии социальной эффективности

ДО	ПОСЛЕ
Подмена оператора ПК приводит к утечке информации	Подмена оператора ПК оперативно определяется по клавиатурному почерку, ПК блокируется
Работодатели следят за эффективностью сотрудников явно, создавая стрессовую психологическую обстановку	Мониторинг клавиатурного почерка происходит в скрытом режиме

4.4.2 Оценка сравнительной эффективности исследования

Для определения эффективности НТИ необходимо рассчитать интегральный показатель финансовой эффективности и интегральный показатель ресурсоэффективности.

Интегральный финансовый показатель разработки определяется как:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{ri}}{\Phi_{\text{max}}}, \quad (13)$$

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

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

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Для исполнения 1: $I_{\text{финр}}=221798,1/240000= 0,92$.

Для исполнения 2: $I_{\text{финр}}=240000/240000= 1$.

Для исполнения 3: $I_{\text{финр}}=231500/240000= 0,96$.

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

$$I_{pi} = \sum a_i \cdot b_i, \quad (14)$$

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

a_i – весовой коэффициент i -го варианта исполнения разработки;

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

n – число параметров сравнения.

В таблице 20 представлена сравнительная оценка характеристик вариантов исполнения проекта.

Таблица 20 – Сравнительная оценка характеристик вариантов исполнения проекта

Оценочные критерии	Весовой коэффициент параметра	Текущий проект	Аналог 1	Аналог 2
1. Функциональная мощность	0,2	4	5	3
2. Скорость работы	0,3	5	3	4
3. Потребность в ресурсах памяти	0,1	4	3	4
4. Зависимость от платформы	0,1	2	2	2
5. Удобство эксплуатации	0,1	5	5	4
6. Надежность	0,2	4	5	4
Итого	1			

Таким образом, $I_{p1}=0,2*4+0,3*5+0,1*4+0,1*2+0,1*5+0,2*4 = 4,2$.

$I_{p2}=0,2*5+0,3*3+0,1*3+0,1*2+0,1*5+0,2*5 = 3,9$.

$I_{p3}=0,2*3+0,3*4+0,1*4+0,1*2+0,1*4+0,2*4 = 3,6$.

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

$$I_{исп.1} = \frac{I_{р-исп1}}{I_{финр}} \quad (15)$$

$$I_{исп1} = 4,2/0,92 = 4,57.$$

$$I_{исп2} = 3,9/1 = 3,9.$$

$$I_{исп3} = 3,6/0,96 = 3,75.$$

После этого необходимо определить сравнительную эффективность исполнений разработки, которая позволит выбрать самый выгодный вариант разработки с позиции финансовой и ресурсной эффективности:

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.2}} \quad (16)$$

$$\mathcal{E}_{cp1} = 4,57/4,57 = 1.$$

$$\mathcal{E}_{cp2} = 3,9/4,57 = 0,85.$$

$$\mathcal{E}_{cp3} = 3,75/4,57 = 0,82.$$

Таблица 21 демонстрирует расчет сравнительной эффективности разработки.

Таблица 21 – Сравнительная эффективность разработки

п/п	Показатели	Текущий проект	Аналог 1	Аналог 2
1	Интегральный финансовый показатель разработки	0,92	1	0,96
2	Интегральный показатель ресурсоэффективности разработки	4,2	3,9	3,6
3	Интегральный показатель эффективности	4,57	3,9	3,75
4	Сравнительная эффективность	1	0,85	0,82

	вариантов исполнения			
--	----------------------	--	--	--

Таким образом, самым эффективным исполнением с позиции ресурсоэффективности и финансовой эффективности является текущий проект. Наименее эффективным является аналог 2, поскольку его интегральный показатель эффективности на 18% ниже, текущего проекта.

5 Социальная ответственность

5.1 Введение

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.

5.2 Производственная безопасность

Опасные и вредные производственные факторы рассмотрены в таблице 22.

Таблица 22 – Опасные и вредные факторы при разработке и эксплуатации системы распознавания пользователя по клавиатурному почерку

Источник фактора, наименование видов работ	Факторы (по ГОСТ 12.0.003-74)		Нормативные документы
	Вредные	Опасные	
1) Разработка системы распознавания клавиатурного	1) Повышенная или пониженная температура воздуха рабочей	1) Опасность поражения электрическим током;	СанПиН 2.2.4.548 – 96 ГОСТ Р 12.1.019-2009

почерка; 2) Эксплуатация системы распознавания пользователя по клавиатурному почерку.	зоны; 2) Повышенный уровень электромагнитных излучений; 3) Недостаточная освещенность рабочей зоны; 4) Повышенная или пониженная подвижность воздуха; 5) Нервно-психические перегрузки.	2) Короткое замыкание; 3) Статическое электричество.	СНиП 23-05-10 ГОСТ Р 12.1.019-2009 ГОСТ 12.1.004-91 СанПиН 2.2.2/2.4.1340-03
--	---	---	---

Повышенная или пониженная температура воздуха рабочей зоны – это вредный фактор физической природы. Слишком высокая или низкая температура воздуха негативно сказывается на внимательности и концентрации сотрудника, ухудшает настроение, а также снижает работоспособность. Пониженная температура воздуха, помимо всего прочего, увеличивается риск возникновения и распространения простудных заболеваний в коллективе.

Повышенная или пониженная температуры воздуха в помещении может наблюдаться при значительном понижении или повышении температуры воздуха за окном. При этом определяющее значение играет работа систем отопления и вентиляции в помещении.

Работа офисного сотрудника сопряжена с незначительным физическим напряжением и производится в положении сидя, поэтому по уровню энергозатрат она относится к категории Ia. Оптимальные значения

температуры воздуха производственных помещений для категории Ia указаны в [30] и представлены в таблице 23.

Таблица 23 – Оптимальные параметры микроклимата на рабочих местах производственных помещений

Период года	Категория работ по уровню энергозатрат, Вт	Температура воздуха, °С
Холодный	Категория Ia (до 139)	23-25
Теплый	Категория Ia (до 139)	20-22

Температура в помещении составляет 24⁰С в холодный период и 22⁰С в теплый период года. Рабочее место соответствует нормам СанПиН 2.2.4.548 – 96.

Повышенный уровень электромагнитных излучений – это физический вредный производственный фактор. Источником излучения в офисном помещении являются системные блоки персональных компьютеров. Электромагнитное излучения негативно воздействует на организм человека, способствует повышению температуры тела и может спровоцировать микропроцессы, изменяющие свойства молекул в организме. В [32] представлены максимально допустимые уровни напряженности электрического и магнитного полей, а также плотности потока энергии электромагнитного поля. Таблица 24 показывает соответствующие значения.

Таблица 24 – ПДУ энергетических экспозиций ЭМП диапазона частот ≥ 30 кГц - 300 ГГц

Диапазоны частот	Предельно допустимые уровни энергетической экспозиции		
	По электрической составляющей, (В/м) ² × ч	По магнитной составляющей, (А/м) ² × ч	По плотности потока энергии (мкВт/см ²) × ч
30 кГц - 3 МГц	20000,0	200,0	-

3 - 30 МГц	7000,0	-	-
30 - 50 МГц	800,0	0,72	-
50 - 300 МГц	800,0	-	-
300 МГц - 300 ГГц	-	-	200,0

Индивидуальные средства защиты сотрудников заключаются в соблюдении безопасной дистанции от глаз до монитора, которая составляет от 0.5 до 1 метра. Коллективные средства защиты – это использование в помещении исключительно жидкокристаллических мониторов, а не устаревших устройств с ЭЛТ.

Недостаточная освещенность рабочей зоны также относится к физическим вредным производственным факторам. Она наблюдается при недостатке ламп освещения, либо сильной пульсации света ламп и других осветительных приборов. Недостаточная освещенность приводит к увеличению напряжения глаз сотрудников, ухудшению зрения, а также снижает уровень работоспособности, бодрости, ухудшает настроение сотрудников.

Поскольку в офисе работа сотрудника происходит с использованием персонального компьютера, необходимо учитывать правила и нормы к освещению, указанные в [33]. Эти нормы представлены в таблице 25.

Таблица 25 – Требования к освещению на рабочих местах, оборудованных ПК (компьютерные залы)

Плоскость освещенности	Вертикальная плоскость освещенности (дисплей компьютера) в КЕО	Горизонтальная плоскость освещенности (рабочий стол) в КЕО
Параметр		
Высота плоскости над полом, м	1,2	0,8

Разряд и подразряд зрительной работы		Б-2	А-2
Искусственное освещение	Освещенность рабочих поверхностей при комбинированном освещении, лк		500/300
	Освещенность рабочих поверхностей при общем освещении, лк	200	400
	Объединенный показатель дискомфорта UGR, не более	-	14
Искусственное освещение	Коэффициент пульсации освещенности, %, не более	-	10
Естественное освещение	КЕО e_n , % при боковом освещении	-	1.2
	КЕО e_n , % при верхнем или комбинированном освещении	-	3.5

Для снижения негативного воздействия недостаточной освещенности на глаза сотрудников предлагается делать гимнастику для глаз каждый час, отвернувшись от экрана.

Офисное помещение имеет следующие размеры: высота 2,5 м, ширина 3 м, длина 6 м. Источниками освещения служат светильники ОД 2-30. Отношение средней освещенности к минимальной Z примем равным 1.1. Коэффициент запаса K определяется исходя из запылённости помещения, в нашем случае $K = 1.5$. Коэффициент отражения от стен $PC=70\%$, коэффициент отражения от потолка $PP=50\%$ в соответствии со СНиП 23-05-95 [33].

Значение η определим по таблице коэффициентов использования различных светильников из [33]. Для этого вычислим индекс помещения по формуле:

$$I = \frac{S}{h \cdot (a+b)}, \quad (17)$$

Площадь помещения составляет 18 м^2 .

$$I = \frac{18}{2,5 \cdot (3+6)} = 0,8$$

Следовательно, коэффициент использования светового потока $\eta = 0,5$. Стандартный световой поток $\Phi_{\text{ст}} = 1060$.

$$E_{\text{ф}} = (N * n * \Phi_{\text{ст}} * \eta) / (S * k * z) \quad (18)$$

$$E_{\text{ф}} = (6 * 2 * 1060 * 0,5) / (18 * 1,5 * 1,1) = 214 \text{ лк}$$

В соответствии с [33] минимальная освещённость рабочих поверхностей в офисных помещениях для работ средней точности при общем освещении должна быть равна 200 лк. Таким образом, освещённость в офисном помещении достаточная и соответствует нормам СНиП.

Повышенная или пониженная подвижность воздуха относится к физическим вредным производственным факторам. Источником является неоптимальная работа систем вентиляции и кондиционирования воздуха. Недостаточная вентиляции приводит к ухудшению качества воздуха, что, в свою очередь, негативно сказывается на работоспособности сотрудников и их внимательности.

Нормы подвижности воздуха, а также влажности воздуха, указаны в [30] и приведены в таблице 26.

Таблица 26 – Нормы подвижности и влажности воздуха в производственных помещениях

Период года	Категория работ по уровню энергозатрат, Вт	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Категория Ia(до 139)	40-60	0,1

Теплый	Категория Ia(до 139)	40-60	0,1
--------	----------------------	-------	-----

Коллективные средства защиты предусматривают оборудование механической вентиляции в помещениях, где невозможна естественная вентиляция. Индивидуальные средства могут включать настольные вентиляторы для сотрудников.

Нервно-психические перегрузки – это вредный производственный фактор психофизиологической природы. Перегрузки возникают в связи с высокой ответственностью и сложностью работы программиста. В результате длительного стресса возможно нарушение функций сердечно-сосудистой и центральной нервной системы. Также в результате перегрузок наблюдается ухудшение работы иммунной системы, что негативно влияет на общее состояние здоровья сотрудника и его работоспособность.

Для защиты от перегрузок сотрудникам рекомендуется:

1. Соблюдать режим работы и отдыха;
2. Поддерживать доброжелательные отношения с коллегами;
3. Заниматься спортом;
4. В тяжелых случаях – проконсультироваться со специалистом и

начать лечение.

Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека – опасный фактор физической природы. При работе с компьютером сотруднику грозит поражение электрическим током в случае короткого замыкания. Это может привести к появлению ожогов, механическим повреждениям тканей, нагреву сосудов, раздражающим воздействиям на ткани.

Требования по электробезопасности регулируются в [34]. Несмотря на то, что офисное помещение относится к категории помещений без повышенной опасности, сотрудникам необходимо соблюдать следующие меры предосторожности при работе с компьютером:

- Не работать на компьютерах, имеющих нарушения целостности корпуса, неисправную индикацию включения питания, нарушения изоляции проводов, а также с признаками электрического напряжения на корпусе;
- Не закладывать провода и шнуры за газовые и водопроводные трубы, за батареи отопительной системы;
- Не выдергивать штепсельную вилку из розетки за шнур. Усилие должно быть приложено к корпусу вилки;
- Не класть на средства вычислительной техники и периферийное оборудование посторонние предметы.

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

В офисном помещении при работе за компьютером может возникнуть чрезвычайная ситуация – пожар. Причинами пожаров могут быть:

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

В [35] представлены требования к системам противопожарной защиты, а также мероприятия по обеспечению пожарной безопасности.

Для защиты от пожаров в офисе необходимо иметь огнетушители, пожарные шкафы и пожарные щиты. Необходим регулярный инструктаж сотрудников по технике безопасности и действиях при пожаре.

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 Правовые и организационные вопросы обеспечения 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.

В ТК РФ[36] и СанПиН 2.2.2/2.4.1340-03[37] устанавливаются следующие правила и вводятся рекомендации, касающиеся работы сотрудников всех организаций на территории РФ:

- Нормальная продолжительность рабочего времени не должна превышать 40 часов в неделю.
- Продолжительность непрерывной работы за компьютером без перерыва не должна превышать 1 час;
- 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;

- 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.

5.6 Вывод

В разделе проанализированы вредные и опасные производственные факторы. Было установлено, что офисное помещение, в котором производилась разработка программы по распознаванию пользователей на основе параметров клавиатурного почерка, соответствует нормам СНиП и СанПиН. Дополнительных средств защиты сотрудников не требуется.

При работе в офисе 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 мало.

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 результате работы были получены новые теоретические сведения об отличительных особенностях клавиатурного почерка, пригодные для использования в качестве дискриминантного признака. В работе рассматривались существующие характеристики КП и был сформирован векторный динамический показатель. Была создана математическая модель клавиатурного нажатия при использовании персонального компьютера.

Составлен алгоритм и реализовано программное приложение для сбора динамических характеристик КП и создания банка данных с образцами почерка пользователей. Полученная таким образом база данных образцов КП содержит образцы КП более 50 пользователей. В ней сохраняются такие сведения как среднее время удержания каждой клавиши, среднее время пауз между нажатиями, количество ошибок ввода, а также уникальные идентификаторы пользователей. Для каждого пользователя сохранено от 1 до 10 образцов почерка.

Кроме того, собранные сведения были дополнены данными из БД европейских пользователей, накопленной учеными из Мадридского автономного университета, Испания. В результате удалось собрать БД из 300 пользователей различного возраста, пола, национальности и рода деятельности.

Был составлен алгоритм, реализующий гибридный метод для решения задачи аутентификации зарегистрированного пользователя или идентификации попытки несанкционированного доступа. Для сравнения образов КП использовался векторный дискриминантный критерий эффективности.

Реализовано программное приложение, имеющее клиент-серверную архитектуру. Серверная часть системы отвечает за хранение и обработку данных о характеристиках КП пользователей. Клиентская часть системы собирает необходимые сведения путем мониторинга клавиатурных нажатий.

Для повышения эффективности алгоритма распознавания планируется исследовать возможность применения биоинспирированных методов анализа. Кроме того, планируется продолжать сбор статистики динамических характеристик клавиатурного почерка и проведение более масштабного тестирования.

Conclusion

As a result of the work, new theoretical knowledge was acquired on the features of keystroke dynamics, suitable for use as a discriminant characteristic. The existing characteristics of keystroke dynamics were considered and a vector dynamic indicator was generated in this work. A mathematical model of keyboard pressing was created for a user with a personal computer.

An algorithm was developed and a software application was implemented to collect the dynamic characteristics of keystroke dynamics and create a data bank with samples of users' keystroke dynamics. The obtained database of keystroke dynamics samples contains data for more than 50 users. It stores such information as the average dwell time of each key, the average flight time, the number of input errors, and unique user IDs. From 1 to 10 keystroke dynamics samples is saved for each user.

Additionally, the collected information was supplemented with data from the database of European users, gained by scientists from the Madrid Autonomous University, Spain. As a result, it was possible to collect a database of 300 users of different ages, gender, nationality and occupation.

An algorithm was implemented. It uses a hybrid method to solve the problem of authenticating a registered user or identifying an attempted unauthorized access. A vector discriminant efficiency criterion was used to compare keystroke dynamics samples.

A software application that has client-server architecture was implemented. The server part of the system is responsible for storing and processing data of the characteristics of user keystroke dynamics. The client part of the system collects the necessary information by monitoring keyboard pressings.

To increase the efficiency of the recognition algorithm, it is planned to investigate the possibility of using bioinspired methods of analysis. In addition, it is planned to continue collecting statistics on the characteristics of keystroke dynamics and to conduct more tests.

Список литературы

1. Технология усиленной аутентификации пользователей информационных процессов / И.А. Ходашинский, М.В. Савчук, И.В. Горбунов, Р.В. Мещеряков. - Управление, вычислительная техника и информатика, 2011 - с.236-248
2. В.А. Ворона. Системы контроля и управления доступом / В.А, Ворона, В.А. Тихонов. – М.: Горячая линия – Телеком, 2010. – 274 с.
3. Banerjee, S. P., Woodard, D. L.: Biometric Authentication and Identification using Keystroke Dynamics: A Survey. Journal of Pattern Recognition Research. 7, 116—139 (2012).
4. P.H. Pisani, A.C. Lorena, A systematic review on keystroke dynamics, J. Braz. Comput. Soc. 19 (4) (2013) 573–587.
5. A. Morales, J. Fierrez, M. Gomez-Barrero, J. Ortega-Garcia, R. Daza, J. V. Monaco, J.Montalvão, J. Canuto, A. George, "KBOC: Keystroke Biometrics OnGoing Competition", Proc. 8th IEEE International Conference on Biometrics: Theory, Applications, and Systems, Buffalo, USA, pp. 1-6, 2016.
6. Junhong Kim, Haedong Kim, Pilsung Kang* Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection // Applied Soft Computing 62 (2018) 1077–1087.
7. Md Liakat Ali · John V. Monaco · Charles C. Tappert · Meikang Qiu Keystroke Biometric Systems for User Authentication // J Sign Process Syst (2017) 86:175–190 DOI 10.1007/s11265-016-1114-9.
8. Teh, P. S., Teoh, A. B. J., & Yue, S., «A Survey of Keystroke Dynamics Biometrics», The Scientific World Journal, 2013, vol. 2013, 24 pages,. doi:10.1155/2013/408280.
9. Анализ причин возникновения ошибок первого и второго рода в системах авторизации основанных на распознавании клавиатурного почерка / А.Н. Савинов. - Программные системы и вычислительные методы – №1(1), 2012 – с. 53-59.

10. Т.В. Жашкова. Нейросетевая идентификация типа личности человека по клавиатурному почерку: статья / Т.В. Жашкова, О.М. Шарунова, Э.Ш. Исянова. – Международный студенческий научный вестник, №3, ч.1, 2015. – 6 с.
11. И.Г. Сидоркина. Три алгоритма управления доступом к КСИИ на основе распознавания клавиатурного почерка оператора: статья / И.Г. Сидоркина, А.Н. Савинов. – Вестник Чувашского университета, №3, 2013. – 9 с.
12. Методы, модели и алгоритмы распознавания клавиатурного почерка в ключевых системах / Савинов А.Н. – 2013.
13. Keystroke Dynamics for User Authentication / Y. Zhong, Y. Deng - 2012.
14. Использование клавиатурного почерка для аутентификации в компьютерных информационных системах / С. В. Скуратов - Безопасность информационных технологий, 2010.
15. Клавиатурная идентификация пользователя на базе метода многосвязного представления / Р. В. Ерусланов - Радиотехнические и инфокоммуникационные системы, 2009.
16. Биометрический метод аутентификации пользователя / В.Ф. Гузик, М.Н. Десятерик - Известия ТРТУ, 2000.
17. D. Gunetti, C. Picardi Keystroke analysis of free text ACM Trans. Inform. Syst. Security, 8 (2005), pp. 312-347.
18. R. Joyce, G. Gupta, Identity authentication based on keystroke latencies, Commun. ACM 33 (2) (1990) 168–176.
19. Y. Zhong and Y. Deng, "A survey on keystroke dynamics biometrics: approaches, advances, and evaluations". In: Y. Zhong, Y. Deng (eds.) Recent Advances in User Authentication Using Keystroke Dynamics Biometrics. Science Gate Publishing, pp. 1-22, 2015.

20. P. Kang, S. Cho, Keystroke dynamics-based user authentication using long and free text strings from various input devices, *Inf. Sci.* 308 (2015) 72–93.

21. Ahmed Mahfouz, Tarek M. Mahmoud, Ahmed Sharaf Eldin A survey on behavioral biometric authentication on smartphones // Research article *Journal of Information Security and Applications*, Volume 37, December 2017, pp. 28-37.

22. O. Alpar Frequency spectrograms for biometric keystroke authentication using neural network based classifier // *Knowledge-Based Syst.*, 116 (2017), pp. 163-171.

23. Adam Goodkind, David Guy Brizan, Andrew Rosenberg, Utilizing overt and latent linguistic structure to improve keystroke-based authentication, *Image and Vision Computing*, Volume 58, 2017, pp. 230-238.

24. Русский алфавит. Частотность букв русского языка (по НКРЯ). Частотность русского алфавита - как часто встречается данная буква в массиве случайного русского текста. [Электронный ресурс] / Инженерный справочник. URL:

<https://www.dpva.ru/Guide/GuideUnitsAlphabets/Alphabets/FrequencyRuLetters/>, дата обращения: 11.03.2018

25. Новые возможности C# [Электронный ресурс] / MSDN. URL: <https://docs.microsoft.com/ru-ru/dotnet/csharp/whats-new/index>, дата обращения: 11.03.2018

26. Текстовый интерфейс пользователя [Электронный ресурс] / Википедия. URL: https://ru.wikipedia.org/wiki/Текстовый_интерфейс_пользователя, дата обращения: 11.03.2018

27. Windows Forms [Электронный ресурс] / MSDN. URL: <https://docs.microsoft.com/ru-ru/dotnet/framework/winforms/>, дата обращения: 11.03.2018

28. Сокеты [Электронный ресурс] / lectures. URL: <http://lecturesnet.readthedocs.io/net/low-level/ipc/socket/intro.html>, дата обращения: 11.03.2018

29. A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos, S. Marcel, "Keystroke Biometrics OnGoing Competition", IEEE Access, In press, pp. 1-11, 2016

30. СанПиН 2.2.4.548 – 96. Гигиенические требования к микроклимату производственных помещений. М.: Минздрав России, 1997. – 14 с.

31. ГОСТ 12.4.011-89. Средства защиты работающих. М.: Минздрав России, 1996. – 7 с.

32. ГОСТ Р 12.1.019-2009 ССБТ Электробезопасность. Общие требования и номенклатура видов защиты. – М.: Стандартинформ, 2010. – 32 с.

33. СНиП 23-05-10. Естественное и искусственное освещение. М.: Минрегион России, 2010. – 76 с.

34. ГОСТ Р 12.1.019-2009 ССБТ Электробезопасность. Общие требования и номенклатура видов защиты. – М.: Стандартинформ, 2010. – 32 с.

35. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования. – М.: Стандартинформ, 2006. – 68 с.

36. Трудовой кодекс Российской Федерации: от 30.12. 2001 г. № 197-ФЗ, введ в действие с 01.02.2002 / Российская Федерация. Законы. — Волгоград ; М. : Изд-во ВолГУ : Либрис, 2002. — 225 с.

37. СанПиН 2.2.2/2.4.1340 – 03. Санитарно-эпидемиологические правила и нормативы. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы. М.: Минздрав России, 2003. – 15 с.

Список публикаций

1. Kochegurova E. A. , Gorokhova E. S. , Mozgaleva A. I. Development of the Keystroke Dynamics Recognition System // Journal of Physics: Conference Series. - 2017 - Vol. 803, Article number 012073. - p. 1-6.
2. Горохова Е. С. , Соколова Е. В. , Демиденко Л. Р. Разработка системы распознавания пользователя по клавиатурному почерку [Электронный ресурс] // Молодежь и современные информационные технологии: сборник трудов XIV Международной научно-практической конференции студентов, аспирантов и молодых ученых. В 2 т., Томск, 7-11 Ноября 2016. - Томск: ТПУ, 2017 - Т. 2 - С. 116-117.
3. Горохова Е. С. Алгоритмы распознавания компьютерного почерка // Молодежь и современные информационные технологии: сборник трудов XIII Международной научно- практической конференции студентов, аспирантов и молодых ученых , Томск, 9-13 Ноября 2015. - Томск: ТПУ, 2016 - Т. 2 - С. 83-84.
4. Горохова Е. С. Программное обеспечение для распознавания клавиатурного почерка // Ресурсоэффективным технологиям - энергию и энтузиазм молодых: сборник научных трудов VII Всероссийской конференции, Томск, 27-29 Апреля 2016. - Томск: Изд-во ТПУ, 2016 - С. 134-135.

Приложение А (Справочное)

Таблица А.1 – Матрица SWOT

	<p>Сильные стороны научно-исследовательского проекта: С1. Простота эксплуатации С2. Низкая стоимость разработки С3. Централизованное хранение данных С4. Низкие требования к аппаратно-программному обеспечению С5. Удобный интерфейс С6. Графическое представление данных</p>	<p>Слабые стороны научно-исследовательского проекта: Сл1. Невысокая точность алгоритма распознавания клавиатурного почерка Сл2. Отсутствие кроссплатформенности Сл3. Длительная разработка</p>
<p>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</p>		
<p>Угрозы: У1. Увеличение конкуренции У2. Прекращение поддержки руководителей проекта У3. Отсутствие интереса к продукту на рынке</p>		

Таблица А.2 – Итоговая таблица SWOT-анализа

	<p>Сильные стороны научно-исследовательского проекта: С1. Простота эксплуатации</p>	<p>Слабые стороны научно-исследовательского проекта: Сл1. Невысокая точность</p>
--	---	--

	<p>C2. Низкая стоимость разработки</p> <p>C3. Централизованное хранение данных</p> <p>C4. Низкие требования к аппаратно-программному обеспечению</p> <p>C5. Удобный интерфейс</p> <p>C6. Графическое представление данных</p>	<p>алгоритма распознавания клавиатурного почерка</p> <p>Сл2. Отсутствие кроссплатформенности</p> <p>Сл3. Длительная разработка</p>
<p>Возможности:</p> <p>V1. Реализация новых функций</p> <p>V2. Повышение отказоустойчивости программы</p> <p>V3. Увеличение спроса на продукт</p> <p>V4. Расширение команды разработчиков для ускорения реализации и поддержки продукта</p> <p>V5. Реализация версий программы для Linux и MacOS</p>	<p>1. V1C2C3C6 – Простота расширения функционала системы.</p> <p>2. V2C2C3C4 – Простота и низкая стоимость изменения каналов связи.</p> <p>3. V3C1C3C4C5C6 – Широкие возможности для увеличения спроса.</p> <p>4. V4C2 – Ускорение разработки.</p> <p>5. V5C2C4 – Расширение рынка сбыта.</p>	<p>1. V1Сл1Сл3 – Необходимость доработки и оптимизации алгоритма.</p> <p>2. V2Сл3 – Модификация приложения требует времени.</p> <p>3. V4Сл1Сл3 – Новые разработчики должны сначала исследовать существующий код и алгоритм.</p> <p>4. V5Сл2Сл3 – Реализация поддержки новой ОС потребует много времени.</p>
<p>Угрозы:</p> <p>У1. Увеличение конкуренции</p> <p>У2. Прекращение поддержки руководителей проекта</p> <p>У3. Отсутствие интереса к продукту на рынке</p>	<p>1. У3С2 – Недостаточно системный подход к разработке ПО</p>	<p>1. У1Сл1Сл2Сл3 – Конкуренты смогут разработать ПО сходного функционала более быстро и качественно.</p> <p>2. У2Сл1Сл3 – Руководитель проекта недоволен реализацией.</p> <p>3. У3Сл1Сл2 – Продукт не удовлетворяет ожидания пользователей.</p>

Таблица А.3 – Бланк оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности	Уровень имеющихся
-------	--------------	-------------------------	-------------------

		научного проекта	знаний у разработчика
1.	Определен имеющийся научно-технический задел	5	5
2.	Определены перспективные направления коммерциализации научно-технического задела	3	3
3.	Определены отрасли и технологии (товары, услуги) для предложения на рынке	4	5
4.	Определена товарная форма научно-технического задела для представления на рынок	5	5
5.	Определены авторы и осуществлена охрана их прав	3	4
6.	Проведена оценка стоимости интеллектуальной собственности	2	2
7.	Проведены маркетинговые исследования рынков сбыта	3	3
8.	Разработан бизнес-план коммерциализации научной разработки	3	3
9.	Определены пути продвижения научной разработки на рынок	4	3
10.	Разработана стратегия (форма) реализации научной разработки	4	4
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	2
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	3	4
13.	Проработаны вопросы финансирования коммерциализации научной разработки	4	4
14.	Имеется команда для коммерциализации научной разработки	5	5
15.	Проработан механизм реализации научного проекта	5	5
	ИТОГО БАЛЛОВ	47	57

Таблица А.4 – Календарный план-график проведения НИОКР

№ работ	Вид работ	Исполнители	T _{кi} кал.дн.	Продолжительность выполнения работ													
				февр.		март			апрель			май			июнь		
				2	3	1	2	3	1	2	3	1	2	3	1	2	
1	Обзор существующего ПО	С	5	■													
2	Обзор существующих методик	Р, С	14	▨	▨												
3	Разработка методики и алгоритма	Р, С	14			▨	▨										
4	Реализация изменений	С	15						■	■	■						
5	Сбор тестовых данных	С	9									■	■				
6	Тестирование продукта	С	5										■				
7	Оценка эффективности полученных результатов	Р, С	9											■	■		
8	Составление пояснительной документации	С	9													■	■

▨ – руководитель

■ – студент

Таблица А.5 – Группировка затрат по статьям

Вид работ	Статьи									
	Сырье, материалы (за вычетом возвратных отходов), покупные изделия и полуфабрикаты	Специальное оборудование для научных (экспериментальных) работ	Основная заработная плата	Дополнительная заработная плата	Отчисления на социальные нужды	Научные и производственные командировки	Оплата работ, выполняемых сторонними и организациями и предприятиями	Прочие прямые расходы	Накладные расходы	Итого плановая себестоимость
1. Разработка системы распознавания клавиатурного почерка	313,12	39500	72567	8708	24382,4	8127,5	3000	0	65200	221798,1

Таблица А.6 – Расчет чистой текущей стоимости по проекту в целом

№	Наименование показателей	Шаг расчета				
		0	1	2	3	4
1.	Выручка от реализации, тыс.руб	0	266	266	266	266
2.	Итого приток, тыс.руб	0	266	266	266	266
3.	Инвестиционные издержки, тыс.руб.	-222	0	0	0	0
4.	Операционные затраты, тыс. руб С+Ам+ФОТ	0	27	27	27	27
5.	Налогооб. прибыль		239	239	239	239
6.	Налоги, тыс. руб Выр-опер=донал.приб*20%		48	48	48	48
7.	Итого отток, тыс.руб. Опер.затр+налоги	-222	75	75	75	75
8.	Чистый денежный поток, тыс.руб. ЧДП=Пчист+Ам Пчист=Пдонал.-налог	-222	195	195	195	195
9.	Коэффициент дисконтирования (приведения при $i = 20\%$)	1	0,833	0,694	0,578	0,482
10.	Дисконтированный чистый денежный поток, тыс.руб. (с8*с9)	-222	162	135	113	94
11.	То же нарастающим итогом, тыс.руб. ($NPV = 223$ тыс.руб.)	-280	-118	17	130	223

Таблица А.7– Зависимость NPV от ставки дисконтирования.

No	Наименование показателя	0	1	2	3	4
1	Чистые денежные потоки	-221	194	194	194	194
2	коэффициент дисконтирования					
	$i=0,1$	1	0,909	0,826	0,751	0,683
	$i=0,2$	1	0,833	0,694	0,578	0,482
	$i=0,3$	1	0,769	0,592	0,455	0,35

	i=0,4	1	0,714	0,51	0,364	0,26	
	i=0,5	1	0,667	0,444	0,295	0,198	
	i=0,6	1	0,625	0,39	0,244	0,095	
	i=0,7	1	0,588	0,335	0,203	0,07	
	i=0,8	1	0,556	0,309	0,171	0,095	
	i=0,9	1	0,526	0,277	0,146	0,077	
	i=1,0	1	0,5	0,25	0,125	0,062	
3	Дисконтированный денежный поток, тыс. руб						
	i=0,1	-221	177	160	146	133	395
	i=0,2	-221	162	135	112	93	282
	i=0,3	-221	149	115	88	68	200
	i=0,4	-221	139	99	70	50	138
	i=0,5	-221	129	86	576	38	90
	i=0,6	-221	121	76	47	18	42
	i=0,7	-221	114	65	39	13	11
	i=0,8	-221	108	60	33	18	-1
	i=0,9	-221	102	53	28	15	-21
	i=0,10	-221	97	48	24	12	0

Приложение Б
(справочное)

Identification of computer systems users on the basis of keystroke dynamics characteristics

Студент

Группа	ФИО	Подпись	Дата
8ВМ6Г	Горохова Екатерина Сергеевна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Кочегурова Елена Алексеевна	к.т.н		

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Кудряшова Александра Владимировна			

1 Overview of technologies and solutions

1.1 Goals and objectives

The goal of this work is to develop recognition theory as well as work out new dynamics characteristics of keystroke for gaining user identification and authorization with the further implementation of software application prototype.

The following objectives must be met in order to achieve the goal.

1. Analysis of trends and approaches to the usage of biometrics for user authentication.
2. Investigating modern approaches to the users keystroke dynamics recognition.
3. Analysis of the existing keystroke dynamics. Evaluating their availability for user recognition. Development of a new vector criteria for comparing two or more examples of keystroke dynamics.
4. Analysis of keystroke dynamics formation mechanism while pressing keys. Developing a math model of keystroke dynamics formation.
5. Development of an algorithm for collecting examples of keystroke dynamics and obtaining vector dynamic criteria. Creation of software for collecting statistics.
6. Collecting statistics of users keystroke dynamics. Creating a database with the keystroke dynamics characteristics of different population groups.
7. Statistics analysis, processing data and finding patterns. Investigating recognition ability of statistic data.
8. Research of existing criteria of difference (proximity) between keystroke dynamics examples. Creating vector discriminant criteria for evaluating proximity of keystroke dynamics, including the frequency of Russian letters and existing proximity metrics.
9. Development of a hybrid keystroke dynamics analysis method for detecting the attempts of unauthorized entries and authentication of registered users.

10. Development of software for collecting keystroke dynamics examples, authorization of registered users and preventing attempts of unauthorized entries.

1.2 Present state of the problem

FRR (False Rejection Rate) and FAR (False Acceptance Rate) are widely used characteristics for analyzing the effectiveness of authentication algorithms.

There are 4 possible cases while an example of keystroke dynamics is compared with a gauge.

- Two examples belong to the same user and the system evaluates them as similar – this is right behavior of the system.
- Two examples belong to different users and the system evaluates them as different – this is also right behavior of the system.
- Two examples belong to the same user but the system evaluates them as different – this is FRR, because the right hypothesis is rejected.
- Two examples belong to different users but the system evaluates them as similar – this is FAR because the wrong hypothesis is accepted.

There are 3 main stages in the operation of the keystroke dynamics user recognition system.

1. Data collection is reading and saving information about keystroke dynamics.
2. Data filtering. The information about service keys is deleted and the information about user's characteristics is distinguished at this stage:
 - a) Amount of mistakes made by user during typing;
 - b) Flight time;
 - c) Dwell time;
 - d) Overlapping of keys presses;
 - e) Rhythm;
 - f) Typing speed.
3. Statistical processing and recording in the database.

In the simplest case all these stages are carried out by a single component on a single computer. Nevertheless, client-server architecture is more preferable in general case, because it allows simultaneous work of many users.

The components may be divided as the following picture shows.

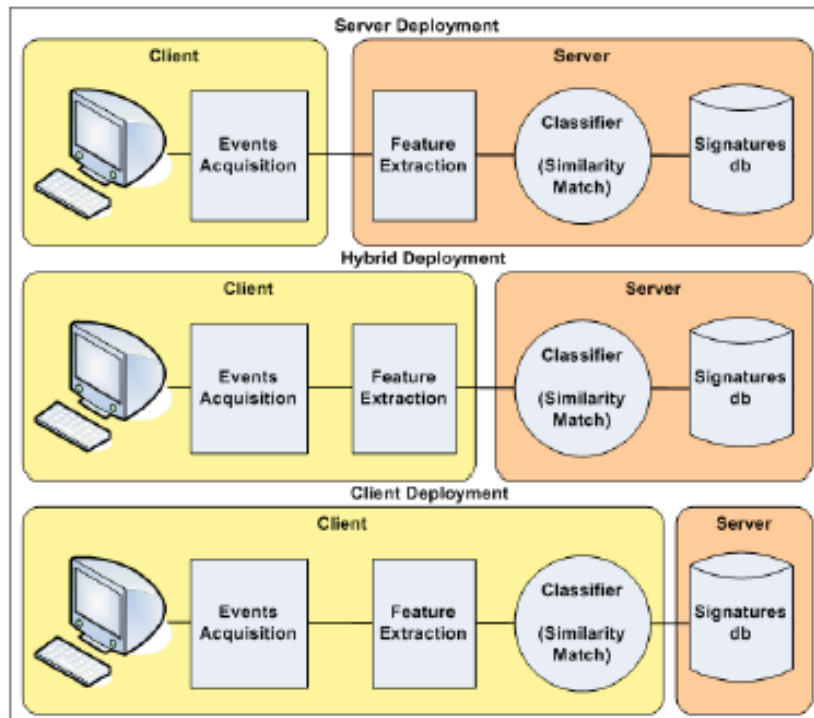


Figure B.1 – Variants of components division

User recognition algorithms using keystroke dynamics may be divided into those which demand a user to input arbitrary or predetermined text, i.e. short password. Algorithms of the first type allow tracking unauthorized substitution of the operator, while algorithms of the second type are easier to implement.

Anyway the system has two operational modes: training and identification.

Stability characteristics of keystroke dynamics are shown in [2] and [3]. The mark means suitability of keystroke dynamics or analysis. Users with unstable keystroke dynamics should be filtered out in order to avoid discrediting of the system.

Table B.1 – Stability characteristics of keystroke dynamics

Errors, %	Arrhythmia, %	Speed, chars/min	Overlapping characteristics	Mark

			Amount of overlapping, %	Amount of finger used	
Less than 2	Less than 10	More than 200	More than 50	All	Excellent
Less than 4	Less than 15	More than 150	More than 30	Majority	Good
Less than 8	Less than 20	More than 100	More than 10	Some	Satisfactory
More than 8	More than 20	Less than 100	Less than 10	Single	Failed

Usually it takes 6 months of working with a keyboard to stabilize keystroke dynamics.

The common difficulties in keystroke dynamics analysis include:

- Variation of keystroke dynamics parameters depends on psycho-physical state of the user;
- Variation of keystroke dynamics parameters depends on the keyboard used. In [2] it is said that the study of the user's keystroke dynamics with different keyboards gave a spread of probability of authentication by 0.5%.
- The necessity of collecting a large amount of statistics for every keystroke dynamics research and lack of existing databases with keystroke dynamics examples.

Users should be classified as known and strangers in the process of authentication. Herewith, the following methods are used [3-10]:

- Geometric methods that calculate the proximity measure between two vectors (Hamming measure, Euclidean measure, etc.);
- Methods based on artificial neural networks;
- Methods of parametric classifier training.

Method of parametric classifier learning supposes calculation of the following characteristics for every sample of keystroke dynamics.

Arrhythmia of the set α and mean value m -pause

$$\bar{m}^{\text{pause}} = \frac{\sum_{i=1}^{n-1} t_i^{\text{pause}}}{(n-1)}; \alpha = \sqrt{\frac{\sum_{i=1}^{n-1} \left(\frac{t_i^{\text{pause}}}{t_{\text{max}}^{\text{pause}}} - \bar{m}^{\text{pause}} \right)^2}{n-2}}. \quad (\text{B.1})$$

Arrhythmia of dwell time β and mean value m -press

$$\bar{m}^{\text{press}} = \frac{\sum_{i=1}^n t_i^{\text{press}}}{n}; \beta = \sqrt{\frac{\sum_{i=1}^n \left(\frac{t_i^{\text{press}}}{t_{\text{max}}^{\text{press}}} - \bar{m}^{\text{press}} \right)^2}{n-1}}. \quad (\text{B.2})$$

Typing speed

$$\bar{s} = \frac{(t_n^{\text{up}} - t_1^{\text{down}})}{60} s_{\text{max}}, \quad (\text{B.3})$$

Where s_{max} is maximum typing speed

The method of parametric classifier has achieved the following results: FRR = 8%, FAR=7%.

The maximum precision of 95% has been achieved using neural network for determination type of personality (extrovert introvert). A microphone has been used in this study in order to record sounds of keystrokes. At the same time, the intensity of the keystrokes has been analyzed.

In case of using a multilayer perceptron to authenticate users by their keystroke dynamics, the results include FAR = 0% and FRR = 1%.

Another promising method is a three-layer Rosenblatt perceptron.

- The first layer is input layer. It contains k formal neurons with linear activation function, where k is the dimension of input vector with keystroke dynamics characteristics.

- The second layer is hidden. It contains k formal neurons with sigma activation function

- The third layer is an output one. It contains p formal neurons with sigma activation function, where p is the number of known users in the system.

Dwell time can be differentiated as keystrokes with overlapping and keystroke without overlapping in order to increase the precision of user recognition algorithms. A graph of the bipolar normal distribution of the dwell time serves as justification.

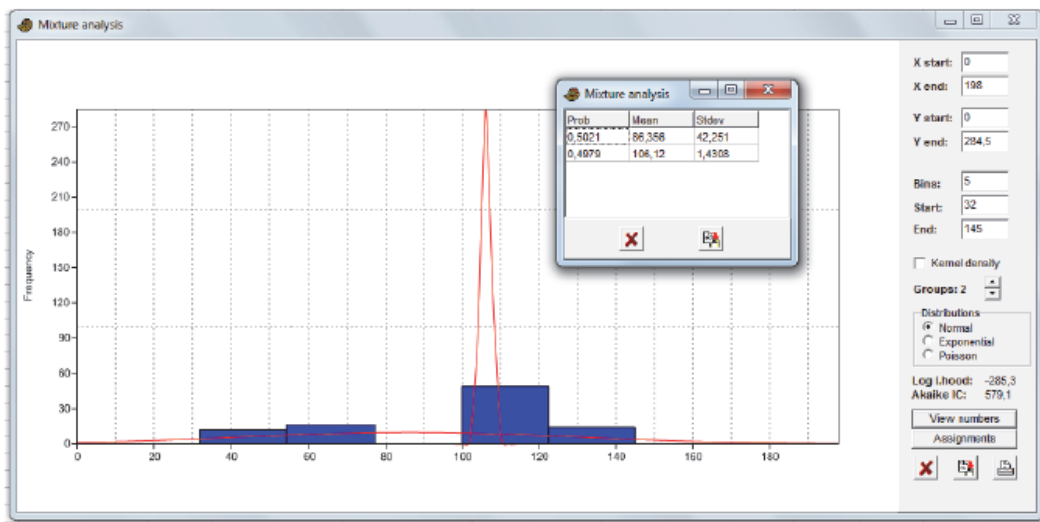


Figure B.2 – Analyses of data set with dwell time using PAST software
The dwell time of key “A” was recorded 91 times during the experiment:

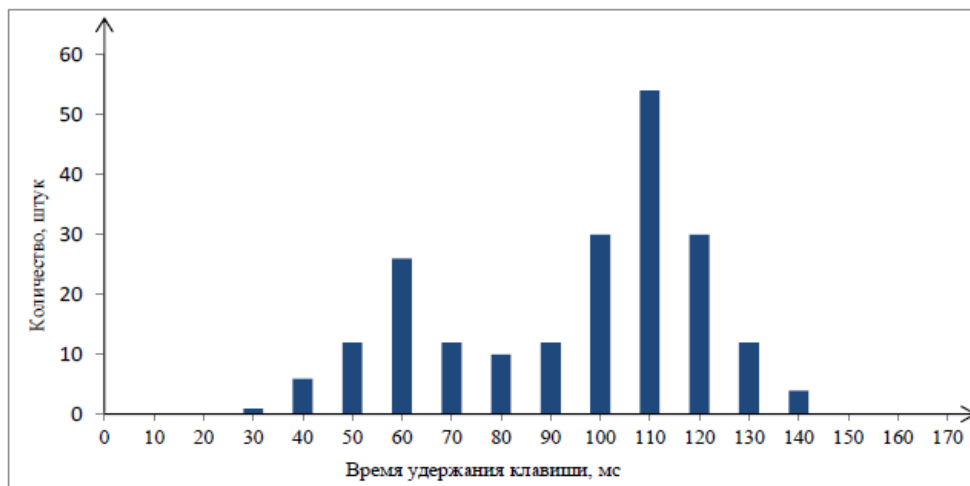


Figure B.3 – Bimodal distribution of dwell time

Euclidean measure, Mahalanobis measure, Manhattan distance can be used for calculating a distance between two vectors of keystroke dynamics. The following classifiers have been used:

- Method of K-means;
- K-nearest neighbors;
- Bayesian classifier;
- Fuzzy logic;
- Neural networks;
- Support vector method.

Hamming measure has been used for the comparison of binary vectors. Hamming distance determines the total number of the differences between the elements of the vector, which stands at the same position. Binary vector is obtained by comparison of a test sample and a standard sample of keystroke dynamics. Furthermore, this work suggests using a trigraph, which is a combination of three consecutive symbols. Trigraph duration is a time interval between pressing the first and the third keys. The best obtained result includes FAR = 0.04% and FRR=6.4%.

The method of multiply connected representation of events when using the keyboard has been used to analyze arbitrary text. A mechanism scheme of classifying is shown in the figure B.4.

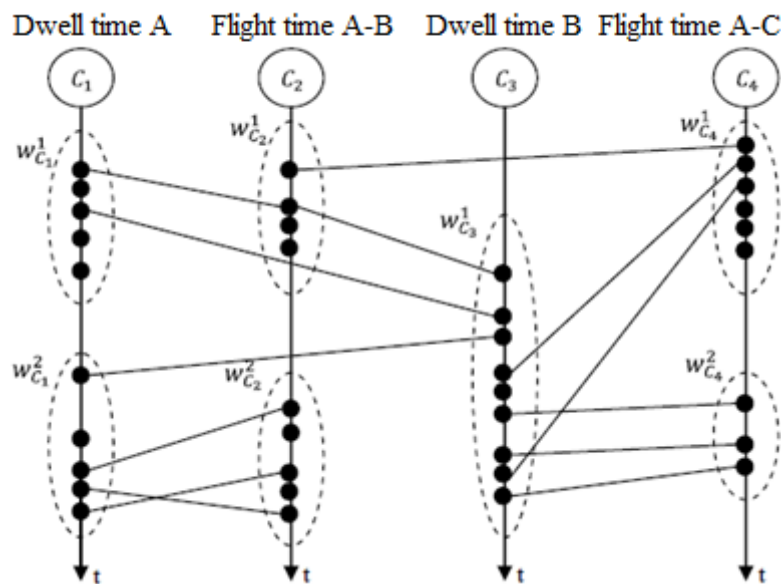


Figure B.4 – A multiply connected representation of keystroke dynamics

During the experiment users were asked to input a 1000 symbols long text. User substitution was successfully determined.

In some cases, the keystroke dynamics authentication problem can be solved by eliminating blunders and then solving the problem of testing the hypothesis of equality of the distribution centers of two normal populations. The results are relevant for input of short pass phrases. The algorithm's precision reaches 97%.

1.3 Practical application

This work is focused on the research of keystroke dynamics and the development of the methodology for keystroke dynamics analysis aimed at user authentication.

Keystroke dynamics analysis helps to solve a range of problems such as:

- Determination of user substitution made on time. For example, imagine that an authenticated user left the workstation and a computer is unlocked. At this time, a stranger can access all the data of the authenticated user. Constant monitoring of keystroke dynamics can determine a user substitution, even if authentication has already been successfully passed.

- Keystroke dynamics can be used as an additional authentication tool. For this purpose dynamic characteristics are analyzed while a user enters their login and password.

- Identification of the user's psycho-physical state. Keystroke dynamics change when a user is overtired. They become less rhythmic, flight time increases and the number of input errors increases as well. Similar problems arise in the state of alcohol and drug intoxication. The presence of staff at work in the described state is counterproductive, and in some cases carries a risk to life and health. Detection of these states is also possible with the help of constant monitoring of keystroke dynamics characteristics.

- Identification of authorship of a documents, program code, etc. This problem is relevant for the systems that allow conducting exams and tests online. In the case of permanent confirmation of the personality with the help of keystroke dynamics, the candidate has the opportunity to take tests from home.

Work [2] shows that up to 140 muscles are involved in the process of typing on a keyboard. Based on the assumption that 20% of the total number of muscles has the most significant effect, we get 28 muscles, and, accordingly, a 28-dimensional control problem. The large dimension problem allows us to talk about the uniqueness of the users' keystroke dynamics.

1.4 Relevance of the research

The relevance of the research is confirmed by the number of publications, which is increasing every year. Work [11] provides the graph of the dependence of the number of publications on a year. The abscissa represents the year, and the ordinate shows the number of publications. The graph is shown in figure B.5.

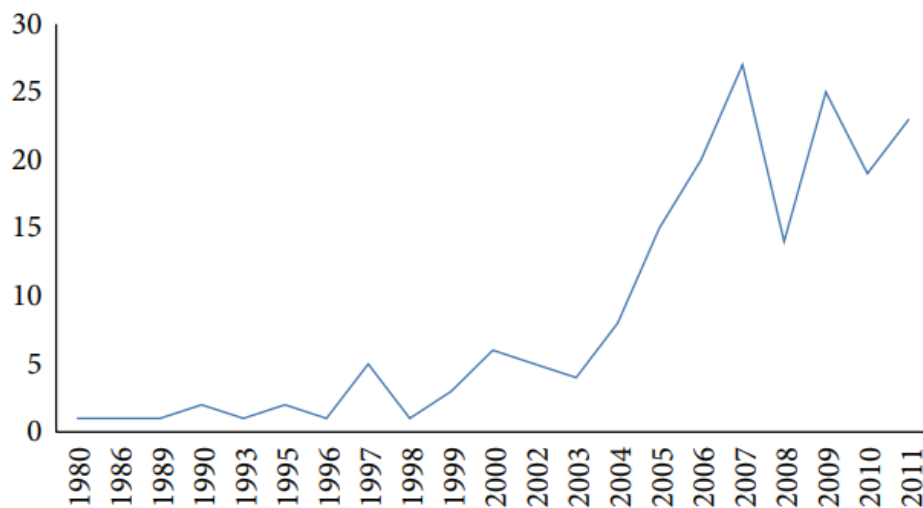


Figure B.5 – Amount of publications on the research topic

Приложение В (Справочное)



EDISON. Центр разработки программного обеспечения
+7 (499) 500-14-94
<http://www.edsd.ru>

АКТ

о принятии к внедрению результатов выпускной квалификационной работы студента Гороховой Екатерины Сергеевны на тему «Идентификация пользователей компьютерных систем на основе динамических характеристик клавиатурного почерка»

Материалы выпускной квалификационной работы Гороховой Е.С. внедрены в практическую деятельность ООО «ЭДИСОН».

В представленных материалах аргументированно проанализирована проблема идентификации пользователя с помощью клавиатурного почерка. Особое внимание уделено разработке методики обучения и распознавания.

Разработанное в результате исследования программное обеспечение внедрено в работу компании в качестве компонента системы «Большой Брат». Его внедрение позволило повысить точность учета трудовой деятельности сотрудников.

Утверждаю:

Зам. директора

ООО «ЭДИСОН»



Строкольская Е.С.

23 мая 2018 года