ЗАКЛЮЧЕНИЕ

экспертной комиссии разового диссертационного совета 6D.KOA-049 при Таджикском техническом университете имени академика М.С. Осими по диссертационной работе **Косимова Абдунаби Абдурауфовича** на тему **«Статистические закономерности распознавания однородности текстов с помощью γ-классификатора»**, представленной на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Экспертная комиссия в составе: председателя доктора физико-математических наук, профессора, академика НАНТ Илолова М.И. (член диссертационного совета) и членов доктора технических наук, профессора Гафарова А.А. (член диссертационного совета) и доктора технических наук, профессора Муминова Б.Б. (член диссертационного совета), сформированная по решению разового диссертационного совета 6D.KOA-049 при Таджикском техническом университете имени М.С. Осими, протокол №2, от «15» марта 2024 года, рассмотрев диссертационную работу **Косимова Абдунаби Абдурауфовича** на тему «**Статистические закономерности распознавания однородности текстов с помощью γ-классификатора**» на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», приняла следующее решение:

Актуальность исследования. Настоящая диссертация является составной частью глобальной научной проблемы – автоматической обработки информации на естественном языке, признанной одной из актуальных проблем современной науки. С надеждами на успешное разрешение последней, связан вопрос о способности современной цивилизации контролировать, упорядочивать, осмысливать и использовать лавинообразный приток знаний, порождаемый её собственной деятельностью.

Одной из граней этой проблемы является проектирование автоматических систем распознавания новизны и адресности информации, охватывающих такие вопросы, как компиляция, плагиат, заимствование, идентификация авторства, сходство произведения и его перевода и другие. В связи с развитием информационных технологий исследования в этой области знания заметно интенсифицировались по всему миру. Многочисленные научные публикации во всех высокоразвитых странах показывают особую роль данной проблематики, её непосредственное влияние на развитие науки и техники, на прогресс в сфере искусственного интеллекта, на широкомасштабные приложения в мировой экономике.

Именно в этом заключается актуальность темы настоящей диссертации, что подтверждается также и постановлением Правительства Республики Таджикистан «Об утверждении программы применения и развития информационных технологий в таджикском языке» от 06.06.2005, № 188, Указом Президента Республики Таджикистан об объявлении 2020-2040 гг. «Двадцатилетием изучения и развития естественных, точных и математических наук в сфере науки и образования» от 31.01.2020, №1445, и поручением, озвученным Президентом Республики Таджикистан, Лидером нации, уважаемым Эмомали Рахмоном в своем ежегодном Послании Маджлиси Оли о принятии и реализации Национальной стратегии развития искусственного интеллекта для разработки и широкого использования современных технологий в различных сферах экономики страны, 21 декабря 2021 года.

Цель исследования заключается в алгоритмизации процесса распознавания однородности текстов и реализации его в виде компьютерного программного комплекса.

Научная специальность диссертационной работы соответствует паспорту научной специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» по следующим пунктам:

1. Модели, методы и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования;

3. Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем;

4. Системы управления базами данных и знаний;

5. Программные системы символьных вычислений;

7. Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения.

Научная новизна диссертации состоит в следующем:

1) исследована информативность нетрадиционных признаков на предмет количественного описания таджикских текстов;

2) установлена статистическая эффективность π математической модели опознавания авторов произведений таджикской классической поэзии (π = 1.00) на основе триграмм, современной поэзии (π = 0.98) с помощью униграмм и современной прозы (π = 0.96) на основе распределения длин предложений (в словах);

3) установлена 100%-ная статистическая эффективность путем применения метрического γ-классификатора и метода ближайшего (по расстоянию) соседа идентифицировать авторов произведений – убывающих по размерам

последовательности текстовых фрагментов от величины в 7000 слов (40000 символов) вплоть до 20 слов (100 символов);

4) для целей существенного сокращения объёма вычислительных процедур установлена возможность эффективного использования не всех, а только высокочастотных элементов цифрового портрета (ЦП) текстов;

5) установлена статистическая эффективность применения γ-классификатора и исследована пригодность ЦП на основе распределения частотности различных алфавитных элементов текста для распознавания других признаков однородности, таких как тематики текста, язык, группа языков, оригинал и его перевод, стиль произведений и шифры научных работ;

6) исследованы статистические закономерности опознавания авторов и языков произведений на корпусах художественных литературных произведений с помощью γ-классификатора;

7) γ-классификатор и метод ближайшего соседа были протестированы на случайных выборках текстов, распознаются с достаточно высокой точностью признаки однородности произведений различных модельных коллекций и корпусов;

8) установлена эффективность применения γ-классификатора для атрибуции искусственно сгенерированных поэм «Шахнаме» А. Фирдоуси по обучению рекуррентных нейронных сетей LSTM (Long short-term memory);

9) исследовано влияние порядка ЦП текста на распознавание однородности произведения с помощью γ-классификатора;

10) впервые в Таджикистане создан объектно-ориентированный компьютерный программный комплекс распознавания (идентификации) однородности текста на основе различных ЦП текстов и γ-классификатора среди сколь угодно большого числа текстов.

Достоверность и обоснованность полученных результатов подтверждены сериями вычислительных экспериментов, в которых посредством γ-классификатора и метода ближайшего соседа распознаются с достаточно высокой точностью самые разные типы «однородностей» в произведениях различных модельных коллекций и корпусов.

Методы исследования. Для решения задач, указанных в рубрике «Цель работы», использовались машинное обучение, кодирование информации, методы математической статистики, вычислительного эксперимента, теории множеств, системного анализа, распознавания и объектно-ориентированного программирования для разработки программных средств.

Положения, выносимые на защиту: экспериментальное доказательство эффективности применения γ-классификатора с помощью различных ЦП текста для распознавания однородности текстовой информации.

Теоретическая значимость работы состоит в том, что в ней эксперимен-

тально опробирован новый метод классификации дискретных случайных величин и установлена эффективность его применения для целей распознавания авторства и для самых разных типов «однородностей» произведений художественной литературы для любых естественных языков на основе различных ЦП текста.

Практическая ценность работы состоит в том, что она нацелена на применение созданного в ней компьютерного программного комплекса *в государственной административной деятельности* для автоматизации процесса обработки текстовой информации, *в сфере криминалистики* для установления авторства анонимных текстов, *в области образования и науки* для обнаружения плагиата в курсовых и дипломных проектах, а также в представленных к защите кандидатских и докторских диссертациях.

Комплекс программ под названием **«THR»** (text homogeneity recognition) применён в следующих организациях:

1. Академия Министерства внутренних дел Республики Таджикистан.
2. Государственный комитет национальной безопасности Республики Таджикистан.
3. Институт языка и литературы имени Рудаки НАНТ.
4. Институт математики имени А.Джураева НАНТ.
5. Таджикский технический университет имени академика М.С. Осими.

Построенный с широким использованием математических моделей и высокого уровня программирования комплекс, в частности, предназначен для развития таджикского языка с использованием возможностей информационных технологий.

Данный комплекс программы является важным как с точки зрения компьютерной лингвистики, так и с точки зрения литературоведения, и направлен на оказание практической помощи исследователям в области языка, литературы, математики и информационных технологий. Среди них он призван определить и распознать стиль каждого автора, особенности отдельных произведений разных авторов, частоту встречаемости букв, слогов, слов, словосочетаний, состав слов в отдельных произведениях, создание различных математических моделей.

Достоверность результатов и надежность спроектированных автоматических систем и программного комплекса подтверждены соответствующими актами о практическом внедрении, документами о выдаче государственного регистрационного номера интеллектуальной продукции в Национальном патентно-информационном центре Министерства экономического развития и торговли Республики Таджикистан. Достоверность данных также подтверждаются отличиями автора в области науки и признанием его заслуг различными организациями и учреждениями республики. Автор удостоен нескольких престижных наград, включая “Лучший педагог” 2023 года среди стран Содружества Независимых Государств, Почетной грамотой Министерства

образования и науки Республики Таджикистан (2023), а также Почетной грамотой и медалью «100-успешных лиц» стран Содружества Независимых Государств (2024).

Указанные достижения соответствуют высоким стандартам и требованиям, установленным Президентом Республики Таджикистан в рамках программы «Двадцатилетие изучения и развития естественных, точных и математических наук в сфере науки и образования» на период с 2020 по 2040 годы.

Публикации по теме диссертации. По материалам диссертационного исследования опубликовано 73 работ, в том числе 34 (14 без соавторства) из которых опубликованы в журналах, рекомендованных ВАК при Президенте РТ и ВАК РФ, 30 статей в международных сборниках статей и журналах, две монографии и два учебных пособия. В патентно-информационном центре при Министерстве экономического развития и торговли Республики Таджикистан получено 5 свидетельств о государственной регистрации информационных ресурсов и интеллектуальной продукции.

Основное содержание диссертации. Диссертация состоит из введения, шести глав, заключения и приложений. Библиографический список включает 397 наименований. Основная часть диссертации изложена на 271 странице. Диссертация содержит 9 рисунков и 107 таблиц.

Оригинальность содержания диссертации составляет 84,67%. Цитирования оформлены корректно. Заимствованного материала, использованного в диссертации без ссылки на автора либо на источник заимствования не обнаружено. Научных работ, выполненных соискателем ученой степени в соавторстве без ссылок на соавторов не выявлено.

Комиссия рекомендует:

1. Тематика и содержание диссертации **Косимова А.А.** соответствуют специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», что позволяет диссертационному совету принять данную диссертацию к защите.

2. Автореферат полностью отражает основное содержание диссертации. Опубликованные работы соответствуют основным положениям диссертации.

3. Количество публикаций в рецензируемых изданиях соответствует установленным требованиям и стандартам Типового положения о диссертационном совете, порядка присвоения учёных степеней и ученых званий.

4. В качестве официальных оппонентов по диссертации комиссия диссертационного совета 6D.KOA-049 рекомендует назначить следующих специалистов:

- доктора технических наук, доцента Пруцкова Александра Викторовича, профессора кафедры «Вычислительная и прикладная математика», Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет», Российская

Федерация;

- доктора физико-математических наук, профессора, Одинаева Раима Назаровича, заведующего кафедрой математического и компьютерного моделирования механико-математического факультета Таджикского национального университета, Республика Таджикистан;

- доктора технических наук, профессора Рахимова Нодира Одиловича, заведующего кафедрой программного обеспечения информационных технологий Ташкентского университета информационных технологий имени Мухаммада ал-Хоразми Республики Узбекистан.

5. В качестве ведущей организации рекомендуется Российско-Таджикский (Славянский) университет.

6. Разрешить размещения информации о предстоящей защите, текста диссертации и автореферата на официальных сайтах ВАК при Президенте Республики Таджикистан и Таджикского технического университета имени академика М.С. Осими.

Председатель комиссии,
д.ф.-м.н., профессор, академик НАНТ М.И. Илолов

Член экспертной комиссии,
д.т.н., профессор А.А. Гафаров

Член экспертной комиссии,
д.т.н., профессор Б.Б. Муминов

Подписи верны:
Ученый секретарь диссертационного
совета 6D.KOA-049, к.т.н.,доцент Ш.М.Султонзода

IMZONI TASDIQLAYMAN TDIU
INSON RESURSLARINI BOSHQARISH
BO'LIMI BOSHLIG'I S. XASANOV
(imzo)