

«Утверждаю»

Директор Института математики им.
А.Джураева Национальной академии
наук Таджикистана, д.ф.-м.н., академик
НАН Таджикистана, профессор

Рахмонов Зарулло Хусенович



«25» 09 2023 г.

Заключение Института математики им. А.Джураева Национальной академии наук Таджикистана

Диссертационная работа Косимова Абдунаби Абдурауфовича на тему «Статистические закономерности распознавания однородности текстов с помощью γ -классификатора (Қонуниятҳои оморӣ шиноҳти якҷинсагии матн бо истифода аз γ -таснифгар)» выполнена в отделе математического моделирования Института математики имени А. Джураева Национальной академии наук Таджикистана.

В период подготовки диссертации и по настоящее время диссертант Косимов Абдунаби Абдурауфович работает в Таджикском техническом университете имени академика М.С. Осими в должности и.о. доцента кафедры «Автоматизированные системы управления».

В 2011 году окончил Худжандский политехнический институт Таджикского технического университета имени академика М.С. Осими по специальности «Программное обеспечение вычислительной техники и автоматизированных систем».

Научный консультант: Усманов Зафар Джураевич, доктор физико-математических наук, академик НАНТ, профессор.

По результатам обсуждения диссертации на тему «Статистические закономерности распознавания однородности текстов с помощью γ -классификатора» принято следующее заключение:

Диссертация Косимова Абдунаби Абдурауфовича на тему: «Статистические закономерности распознавания однородности текстов с помощью γ -классификатора» посвящена изучению проблемы распознавания однородности текстовых фрагментов.

Анализ текста используется во многих прикладных задачах: поиске текстовых аналогов (в структурированном тексте, например, в описаниях патентов или нормативных документах), извлечении из него знаний и данных, определении однородности текста и др.

В диссертации Косимова Абдунаби Абдурауфовича, посвященной выявлению параметров текста, которые можно использовать при решении задачи определения однородности текста, и определению их эффективности. 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.

Актуальность обозначенной научной проблемы подтверждается теоретическими и практическими работами таджикских и зарубежных исследователей. Теоретическая значимость проблемы связана с изучением комплекса вопросов формирования и исследования пригодности ЦП на основе распределения частотности различных алфавитных элементов текста для распознавания новизны, компиляции, плагиата, заимствования, идентификации авторства и шифров научных работ. Актуальность подобных работ связана с определением особых характеристик текста, которые, не будучи подконтрольны своим создателям, содержат в себе косвенную информацию об авторском стиле и даже индивидуальных качествах автора. Практическая значимость проблемы имеет отношение к государственной административной деятельности, в которой на передний план выдвигается автоматическая обработка текстовой информации; к криминалистике, заинтересованной в установлении преступника по составу преступления и авторов анонимных текстов; к сфере образования и науки, в которых и студенческая молодежь и псевдонаучные работники не прочь воспользоваться компиляцией, заимствованиями, плагиатом при выполнении курсовых и дипломных проектов, представлении к защите кандидатских и докторских диссертаций.

Личное участие автора в получении результатов, изложенных в диссертации:

1. сформировал две электронные коллекции текстов, из которых первая предназначена для предварительного тестирования, а вторая – для оценки перспективности применения γ -классификатора;
2. исследовал цифровой портрет текста (ЦПТ) для распознавания автора текста;
3. установил статистическую эффективность применения γ -классификатора для распознавания авторов произведений;
4. определил минимальный размер незнакомого текста, пригодного для распознавания его автора;
5. исследовал эффективность применения высокочастотных элементов ЦПТ для идентификации автора текста;
6. установил статистическую эффективность применения γ -классификатора и исследования пригодности 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, шифры научных работ и т.д.;
7. исследовал статистические закономерности распознавания однородных текстов на корпусах художественных литературных произведений;
8. определил эффективность применения γ -классификатора для атрибуции искусственно сгенерированных произведений авторов;
9. исследовал влияние порядка ЦПТ на распознавание однородности произведения с

помощью γ -классификатора;

10. спроектировал и реализовал компьютерный программный комплекс для распознавания (идентификации) однородности текста на основе различных ЦП текста и γ -классификатора.

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. ТГУ имени академика М.С. Осими.

Основное содержание диссертационного исследования достаточно полно отражено в 73 работах соискателя, в том числе в 34 работах, опубликованных в журналах, рекомендованных Высшей аттестационной комиссией при Президенте Республики Таджикистан:

Основные результаты диссертации опубликованы в следующих статьях:

1. Косимов, А.А. Оценка эффективности использования униграмм при идентификации текста [Текст] / А.А. Косимов // Доклады Академии наук Республики Таджикистан. – 2017. – Т.60. – № 3-4. – С. 132-137.

2. Косимов, А.А. Оценка эффективности использования триграмм при идентификации текста [Текст] / А.А. Косимов // Известия Академии наук Республики Таджикистан. Отделение физико-математических, химических, геологических и технических наук. – 2017. – №1(166). – С. 51-57.

3. Косимов, А.А. Определение минимального объёма выборки слов для идентификации текста [Текст] / А.А. Косимов // Вестник Таджикского национального университета, Серия естественных наук, Душанбе. – 2017. – №1/5. – С. 178-180.

4. Косимов, А.А. Программный комплекс Tajik_Text_Author [Текст] / А.А. Косимов // Политехнический вестник, Серия: интеллект, инновации, инвестиции. – 2019. – 3(47). – С. 22-28.

5. Косимов, А.А. К вопросу об автоматическом распознавании авторства и стилей произведений таджикско-персидской художественной литературы [Текст] / З.Д. Усманов, А.А. Косимов // Доклады Академии наук Республики Таджикистан. – 2020. – Т.63. – № 1-2. – С. 49-54.

6. Косимов, А.А. Об однородности оригинала и его перевода [Текст] / А.А. Косимов // Доклады Национальной академии наук Таджикистана. – 2021. – Т.64. – № 11-12. – С. 660-665.

7. Косимов, А.А. О распознавании автора текстового фрагмента на основе частотности слогов [Текст] / А.А. Косимов // Политехнический вестник, Серия: интеллект, инновации, инвестиции. – 2021. – 4(56). – С. 59-64.

8. Косимов, А.А. О распознавании автора текстового фрагмента на основе

частотности буквенных биграмм [Текст] / А.А. Косимов // Системы анализа и обработки данных. – Том 85. – № 1. – 2022. – С. 73-82. DOI: 10.17212/2782-2001-2022-1-73-82.

9. Косимов, А.А. О влиянии порядка буквенных униграмм на распознавание автора произведения [Текст] / А.А. Косимов // Доклады Национальной академии наук Таджикистана. – 2022. – Т.65. – № 5-6. – С. 324-330.

10. Косимов, А.А. О влиянии порядка буквенных триграмм на распознавание автора произведения [Текст] / А.А. Косимов // Вестник Филиала МГУ имени М.В. Ломоносова в городе Душанбе, Серия естественных наук. – 2022. – № 1. – С. 14-21.

11. Косимов, А.А. Определение шифра специальности с помощью символьных униграмм [Текст] / А.А. Косимов // Вестник Филиала МГУ имени М.В. Ломоносова в городе Душанбе, Серия естественных наук. – 2023. – Том 1. – №1 (29). – С. 16-24.

12. Косимов, А.А. О влиянии порядка символьных триграмм на определение языка произведения [Текст] / А.А. Косимов // Политехнический вестник, Серия: интеллект, инновации, инвестиции. – 2023. – 1(61). – С. 34-37.

Диссертация Косимова Абдунаби Абдурауфовича на тему «Статистические закономерности распознавания однородности текстов с помощью γ -классификатора» соответствует требованиям Высшей аттестационной комиссии при Президенте Республики Таджикистан, предъявляемым к диссертационным работам, и рекомендуется к защите на соискание учёной степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Заключение принято на заседании общеинститутского семинара Института математики им. А.Джураева НАН Таджикистана. Присутствовало на заседании – 17 чел. Результаты голосования: «за» – 17 чел., «против» – 0 чел., «воздержалось» – 0 чел., протокол №13 от «13» сентября 2023 г. Заключение общеинститутского семинара обсуждено и утверждено на заседание Учёного Совета Института от «22» сентября 2023г., протокол №3.

Председатель заседания, главный научный
сотрудник отдела алгебры, теории чисел и топологии,
доктор физико-математических наук

У.Х.Каримов

Ученый секретарь института,
кандидат физико-математических наук

Н.Н.Назрублов

Начальник ОК Института математики
им. А.Джураева, НАН Таджикистана

М.Р.Маллаева

