

УТВЕРЖДАЮ



Ректор Таджикского  
Национального университета  
доц. профессор, член-корр. НАНТ  
Насриддинзода Э.С.

06 \_\_\_\_\_ 2024 г.

## ОТЗЫВ

**ведущей организации на диссертационную работу  
Худойбердиева Хуршеда Атохоновича  
на тему «Проектирование и реализация автоматических систем  
обработки информации на таджикском языке», представленную к  
защите на соискание ученой степени доктора технических наук по  
специальности 05.13.11 – «Математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей»**

*Актуальность избранной темы исследования.* Диссертация Худойбердиева Хуршеда Атохоновича посвящена важнейшей для современной проблемы разработки и реализации систем автоматической обработки информации на таджикском языке.

Задача программного обеспечения для внедрения инновационных технологий в таджикском языке получившая отражение во многих государственных программных документах, в том числе в Государственной стратегии «Информационно-коммуникационные технологии для развития Республики Таджикистан», Стратегии инновационного развития Республики Таджикистан на период до 2020 года, Национальной Стратегии развития Республики Таджикистан на период до 2030 года, направлена на формирование и развитие электронных словарей, синтеза речи, автоматической проверки орфографии и машинного перевода текста.

Результаты исследования направлены на реализацию задач поставленных в Постановлении Правительства Республики Таджикистан о «Государственной программе внедрения информационно-коммуникационных технологий в общеобразовательных учреждениях Республики Таджикистан», Постановления Правительства Республики Таджикистан о «Концепции формирования электронного правительства в Республике Таджикистан» и Постановления Правительства Республики Таджикистан «Государственный стандарт размещения таджикского алфавита на компьютерной клавиатуре», что убедительно доказывает своевременность и актуальность темы исследования.

Развитие на сегодняшний день новой научной сферы, то есть компьютерной лингвистики, действительно, является актуальным научным звеном на пути создания комплексов программных обеспечений управления контентом на естественном языке и поэтому должна стать самостоятельным объектом исследования таджикского языка.

**Оригинальность диссертации** Худойбердиева Х.А. заключается в комплексном исследовании проблем, связанных с обработкой текстовой информации на таджикском языке и ее преобразования в цифровом формате. Это дает основание утверждать, что научная проблема, сформулированная в диссертации, является актуальной. Ее изучение позволяет разработать математические и программные обеспечения по совершенствованию использования таджикского языка пользователями информационных и компьютерных технологий в контексте совершенствования государственного языка. Поиск новых объектно-ориентированных подходов к исследованию задач и реализации механизма повышения качества контента электронной информации на таджикском языке является необходимым с научной и научно-практической точек зрения.

**Степень обоснованности научных положений,** выводов и рекомендаций, сформулированных в диссертации, можно оценить как достаточно высокую, что подтверждено методами исследования и решением автором поставленных им задач. Оценка степени обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, определяется многообразием и представительностью используемых автором исходных данных, полученных из отечественных научных источников. Выводы и рекомендации, сформулированные автором, базируются на глубокой проработке научных результатов, полученных в ходе оценки публикаций зарубежных и отечественных ученых по различным аспектам определения эффективности развития информационных систем через повышение качества математического обеспечения с поддержкой современных компьютерных технологий.

Автором были исследованы проблемы и особенности проектирования и реализации систем обработки информации в условиях развития информационных технологий. Высокий уровень объективности выводов подтверждается использованием в работе публикаций в периодических изданиях и исследований по компьютерной лингвистике, о чем свидетельствует перечень использованной литературы и ссылки на авторов в основной части диссертации. Аналитическое исследование в научной работе и реализация разработанных систем подтверждаются необходимыми статистическими протестированными данными.

Диссертация Худойбердиева Х.А. является завершённой научно-исследовательской работой, которая отражает структурно и по содержанию достаточно полный комплекс решения проблем обработки текстовой информации на таджикском языке, а именно автоматической проверки правописания, машинный перевод и синтез речи. Автореферат и диссертация написаны в соответствии с требованиями ВАК при Президенте Республики

Таджикистан, предъявляемыми к докторским диссертациям, отражают завершённость проведённого исследования и реализации полученных результатов в комплексе программ TajLINGVO.

В диссертации сочетаются теоретические математические модели и методы, преобразованные в программные алгоритмы. В ней обобщены механизмы сбора, обработки и получения статистических показателей, характеризующие цифровой портрет текста на естественном языке и сравнительный анализ на таджикском языке. Аналитические данные представляют результаты исследования за период с 2007 по 2023 гг., что позволило автору получить соответствующий механизм обработки информации, в том числе методы разработки системы автоматической проверки правописания, автоматического перевода и синтеза речи по заданному тексту. Следует отметить четкую последовательность изложения проблемы теоретических вопросов и их совершенствования в математических и программных обеспечениях посредством разработки автоматических систем обработки информации на таджикском языке.

*Достоверность теоретических результатов работы* подтверждается математической и статистической обработкой информации с применением объектно-ориентированного метода исследования. Выводы, полученные в результате применения математических и программных моделей, согласуются с теоретическими взглядами известных таджикских ученых, исследующих данные проблемы. Кроме того, результаты исследования, выдвинутые на защиту, согласуются с достаточно известными исследованиями, нашедших отражение в работах отечественных и зарубежных ученых. Диссертация Худойбердиева Х.А. построена таким образом, что актуальность исследования раскрывает цель и задачи, дает возможность расширить возможность использования таджикского языка в информационных технологиях.

В ходе исследования автор решает следующие задачи:

- разработка методологии и теоретической концепции автоматической обработки текстовой информации на таджикском языке как объекта научного исследования для определения понятий и теоретических терминов в компьютерной лингвистике;
- разработка методов поиска текстовой информации для анализа экспериментальных данных и их применения в научно-практических исследованиях, электронных словарях и компьютерных тезаурусах на таджикском языке;
- разработка модели предоставления текстовой информации и комплекса алгоритмов реализации автоматического синтеза речи на таджикском языке;
- разработка методов извлечения, представления и обработки данных с целью формирования отдельных элементов текста для реализации автоматической проверки правописания текста на таджикском языке;

- разработка моделей, методов и алгоритмов предварительной обработки данных для решения задачи автоматического перевода текста с таджикского языка на русский язык;
- разработка программного комплекса для реализации всех методов, моделей и алгоритмов обработки информации на таджикском языке;
- проведение экспериментального исследования эффективности систем автоматической обработки информации.

Для решения поставленных задач автором были использованы современные методы систематического анализа, математической статистики, основы представления и обработки наборов данных, а также применены теории алгоритмов, математическое и компьютерное моделирование, анализ и синтез данных, компьютерное моделирование информационных систем, технологии объектно-ориентированного программирования и обработки данных.

Введение отражает актуальность диссертационного исследования, степень разработанности проблемы, цель, задачи, область исследования, метод исследования, научную новизну, теоретическую и практическую значимость полученных результатов, сведения об апробации результатов работы.

В первой главе рассмотрены вопросы, связанные с определением существенных особенностей проектирования и проблем реализации систем автоматической обработки информации на таджикском языке.

Во второй главе дается описание методов автоматической обработки информации на естественном языке, а также их решение с использованием математического моделирования, методов анализа и синтеза информации на таджикском языке.

В третьей главе исследуется компьютерное моделирование разработки систем автоматической обработки информации на естественном языке с учетом объектно-ориентированного подхода с применением языка UML. Разработан проект информационной системы обработки текстовых данных на таджикском языке с учетом модели действия, взаимодействия, структуры и реальности.

Четвертая глава посвящена вопросам проектирования, обработки и реализации системы автоматической проверки орфографии текста на таджикском языке.

В пятой главе дается описание проектирования, разработки и внедрения таджикского автоматического переводчика. Здесь рассмотрены задачи реализации машинного перевода текста с таджикского языка на русский.

Шестая глава посвящена задачам математического моделирования и компьютерной реализации синтеза таджикской речи на основе предложенного текста. Здесь рассмотрены методы конкатенации фонем на примере слогов таджикского языка, предложены подходы и методы

последующего проектирования систем распознавания речи на таджикском языке.

В заключении диссертационного исследования обобщены результаты, сформулированы основные выводы и рекомендации, полученные в ходе исследования. В приложении приведены свидетельство о регистрации программного продукта, выданное Национальным патентным центром Министерства экономического развития и торговли Республики Таджикистан, акты внедрения результатов диссертационной работы.

По результатам диссертационной работы соискателем опубликовано 68 научных работ, в том числе 25 статей (11 без соавторства) в журналах, рекомендованных ВАК при Президенте Республики Таджикистан и ВАК Российской Федерации, 27 статей в международных сборниках статей и журналов. Получено 18 свидетельств о государственной регистрации информационных ресурсов и интеллектуального продукта.

Текст диссертации состоит из 328 компьютерных страниц. К написанию диссертации привлечено значительное количество нормативных источников и научных работ (список источников содержит 322 наименования).

Автореферат и опубликованные научные труды соискателя в полной мере отражают содержание диссертации и все основные положения, выносимые на защиту, характеризующие научную новизну.

Результаты исследования нашли научно-методическое отражение и практическое применение в деятельности ряда организаций и учреждений. Полученные в диссертационной работе результаты внедрены в деятельность отдельных организаций, таких как Худжандский научный центр НАНТ, Управление по инвестициям и управлению государственным имуществом Согдийской области, ГОУ Худжандский государственный университет имени академика Б. Гафурова, кафедра таджикского языка Таджикского государственного университета права, бизнеса и политики, Политехнический институт Таджикского технического университета имени академика М.С. Осими в городе Худжанде, а также ЗАО «Душанбе Сити Банк».

Следует отметить, что ссылки автора на используемый материал сделаны корректно, работа оформлена в соответствии с требованиями. Приведен словарь основных терминов компьютерной лингвистики на таджикском, русском и английском языках, составленный автором в период проведения этапов исследования. Работа содержит два приложения, включающих копии свидетельств о государственной регистрации информационных ресурсов, интеллектуальных продуктов и копии актов о внедрении результатов исследования, представленных в диссертации.

Значимость работы усиливается наличием авторских разработок, информационных ресурсов и их практической реализацией. Полученные автором научные результаты вносят значительный вклад в проектирование и разработку систем автоматической проверки правописания, машинного перевода и синтеза речи на таджикском языке.

Сформулированные в рамках работы практические рекомендации могут быть использованы при проектировании и разработки информационных систем обработки информации и реализации ряд задач компьютерной лингвистики, что позволило автору прийти к выводам, определяющим научную новизну исследования, заключающуюся в разработке программного комплекса для эффективного развития использования таджикского языка в электронном делопроизводстве в организациях.

К наиболее значимым результатам, содержащим научную новизну, относятся следующие:

1. В диссертационной работе автор обобщил теоретические разработки в области государственной поддержки развития информационных технологий, предложил результат авторского анализа состояния компьютерной лингвистики в Республике Таджикистан, который получил развитие начиная с 2000 годов под руководством Академика НАНТ доктора физико-математических наук, профессора Усманова З.Дж. Автор сформировал математическую модель системы обработки текстовой информации на естественном языке и механизм её применения для проектирования ряд информационных систем (глава 1).

2. Исследуя мировую практику математического моделирования, автор определил математические методы обработки информации, которые на последующих этапах преобразовал в компьютерные методы и алгоритмы. Анализируя методы и функции обработки информации, автор предлагает свои методы обработки системы проверки правописания текстовых данных, алгоритмы и методы применения машинного перевода, методы и алгоритмы текстового синтеза речи (глава 2).

3. Диссертационное исследование содержит оценку эффективного метода моделирования информационных процессов и обработки информационных систем текстовых данных на основе возможностей языка моделирования UML. Используются современные средства и CASE-инструменты, такие как MS Visual Studio .Net, IBM Rational Rose, Enterprise Architect. Проанализирована проблема моделирования работы информационной системы с учетом вариантов использования, структурного и объектно-ориентированного моделирования, взаимодействий объектов и физическое моделирование процесса обработки текстовых данных (глава 3).

4. Автором применительно к поставленной цели исследования проведена оценка использования кодировки символов таджикского алфавита в стандарте UNICODE. Разработана структура компьютерного тезауруса таджикского языка, подготовлен интерактивный электронный словарь MultiGANJ. Разработан алгоритм преобразования символов текста в государственный стандарт раскладки букв таджикского алфавита. На основе моделей и математических методов разработаны алгоритм преобразования текста в стандартный алфавит, алгоритм обнаружения орфографических ошибок, алгоритм исправления ошибок, алгоритм проверки орфографии (глава 4).

5. Несомненный интерес представляет предложенные автором модели, методы и алгоритмы машинного перевода текста на таджикском языке на русский и английские языки. Выявлены и проанализированы математические модели и методы, используемые в популярной системе машинного перевода онлайн-переводчика Google. Проведен анализ возможных ошибок, возникающих в процессе обработки машинного перевода, как орфографических, так и стилистических. В переводимом тексте большая вероятность встречаемости собственных имен. Для решения этой проблемы на основе методов статистического и машинного обучения предложена система автоматической транслитерации. Разработаны параллельные корпуса Taj-Rus-Corp и Taj-Eng-Corp. Проект машинного перевода доступен в сети Интернет по адресу [www.tarjumon.tajlingvo.tj](http://www.tarjumon.tajlingvo.tj) (глава 5).

6. Самостоятельное научно-практическое значение имеет система синтеза речи по тексту на таджикском языке. Для программного обеспечения синтеза речи на основе математических моделей и специальных компьютерных методов разработаны такие алгоритмы, как алгоритм озвучивания слов; алгоритм озвучивания цифр и символов; алгоритмы безударного и ударного озвучивания текста; алгоритм озвучивания морфемы слова; алгоритм озвучивания таджикского текста, содержащего русские слова. На основе полученных результатов разработана автоматическая система с возможностью синтеза речи на таджикском языке Tajik text-to-speech и Tajik Text Narrator. Автором предлагается фундаментальные методы синтеза речи на таджикском языке на основе конкатенации слогов, которые в будущем могут быть использованы в качестве основы для решения проблемы автоматического распознавания речи на таджикском языке (глава 6).

Позиция автора по указанным вопросам является обоснованной и значимой для дальнейшего исследования научных подходов к проектированию и реализации систем автоматической обработки информации на таджикском языке. Содержание и выводы диссертационного исследования, и практические рекомендации позволяют сделать вывод, что оно представляет собой завершённую работу, в которой в равной степени представлены теоретические положения проектирования систем автоматической обработки информации на естественном языке.

***Теоретическая и практическая значимость исследования.*** Теоретическая значимость результатов проведенного исследования заключается в том, что сформулированные в работе положения и выводы способствуют развитию теории государственного регулирования инновационной деятельности, пространственной экономики. Практическая значимость проведенного исследования заключается в том, что результаты изысканий, направленные на изучение теоретических и научно методических положений инфраструктурного обеспечения поддержки инновационной деятельности регионов, представлены в виде конкретных предложений и рекомендаций, которые могут быть использованы органами государственного управления. Результаты диссертационного исследования могут быть использованы в научной и практической деятельности.

**Надежность полученных результатов исследования** подтверждается актами о практическом внедрении, а достоверность результатов подтверждается премия имени академика С.У. Умарова в области физико-математических, химических, геологических и технических наук НАНТ, 2015г.; Госпремия для учёных и преподавателей естественных, точных и математических дисциплин, 2021г; диплом 3-й степени республиканского конкурса «Наука – цвет процветания», инновация и нововведение, 2021г.; Почетная грамота и медаль «100 НОВЫХ ЛИЦ» стран содружество независимых государств, 2022 г.

**Личный вклад.** Личный вклад Худойбердиева Х.А. состоит в том, что представленные в диссертации новые научные модели, методы и алгоритмы позволяют решить ряд задач в области компьютерной лингвистики, расширяет радиус научного знания о проектировании процессов обработки информации. Несомненно, заслуживает положительной оценки личный вклад автора в научно-практическом плане, который состоит в разработке автоматических систем проверки правописания, машинного перевода и синтеза речи по тексту на таджикском языке.

**Дискуссионные положения и замечания, отмеченные ведущей организацией.** Данная диссертационная работа исследует комплексные взаимодействия математических моделей, проектирования программных обеспечений, государственных стандартов и правил таджикского языка в процессе разработке систем автоматической обработки информации на таджикском языке. В целом положительно оценивая представленные в диссертации результаты проведённого исследования, следует отметить, что работа не лишена дискуссионных моментов и замечаний, заслуживающих уточнения и дополнительной аргументации, что вполне закономерно для научных исследований:

1. Автором не проводится четкое определение понятий «цифровой портрет» и «объектная модель» текста, что несколько снижает степень восприятия предложенных систем обработки информации на таджикском языке.

2. На наш взгляд, автору целесообразно было бы больше внимания уделить системам искусственного интеллекта, указать ее место в обработке текстовой информации и современных методов ее развития.

3. Предложенную автором в третьей главе методику проектирования информационных систем следовало бы апробировать не только для систем, представленных в четвертом, пятом и шестом главах. Следовало бы проектировать системы обработки информации с общим взаимодействием математических моделей и методов, полученных во второй главе.

4. В работе желательно было бы использовать сравнительный анализ к рассмотрению проблемы синтеза речи на опыте иранских ученых, которые достигли значимых результатов в области голосовых технологий.

Данные замечания носят частный характер и не снижают теоретическую и практическую значимость исследования.



### *Соответствие диссертации и автореферата требованиям*

Положения о порядке присуждения ученых степеней. Диссертация Худойбердиева Хуршеда Атохоновича на тему «Проектирование и реализация автоматических систем обработки информации на таджикском языке», представленная на соискание учёной степени доктора технических наук по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, может быть охарактеризована как законченное научное исследование в области компьютерной лингвистики посредством проектирования и реализации ряда систем автоматической обработки информации на таджикском языке, соответствующее требованиям Положения о присуждении учёных степеней, утвержденного Постановлением Правительства Республики Таджикистан от 30 июня 2021 № 267 года (в редакции пост. Правительства РТ от 26.06.2023 г. № 295). Полученные соискателем результаты отвечают критериям научной новизны, теоретически и практически значимы, являются личным вкладом автора в решение исследуемой проблемы.

Диссертация может быть оценена как научно-квалификационная работа, в которой содержатся математические модели, компьютерные методы обработки информации, алгоритмы и комплекс программных обеспечений для повышения эффективного использования информационных технологий на таджикском языке и развитию компьютерной лингвистики.

Автореферат и опубликованные научные работы дают всестороннее представление о содержании диссертационной работы; содержат всю необходимую информацию, характеризующую полученные в процессе исследования результаты, основные положения и выводы диссертации. Основные выводы и рекомендации диссертационного исследования могут использоваться в моделировании процессов обработки текстовой информации на естественном языке и повышению эффективности делопроизводства в организациях на государственном языке.

Диссертационное исследование выполнено в соответствии с паспортом специальности 05.13.11 - «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» по 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 паспорта специальности.

На основании вышесказанного считаем, что диссертационное исследование Худойбердиева Х.А. соответствует всем требованиям ВАК при Президенте Республики Таджикистан, предъявляемым к диссертациям,

представленным к защите на соискание учёной степени доктора технических наук, а её автор Худойбердиев Х.А. заслуживает присуждения ему учёной степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Худойбердиев Х.А. выступил с докладом по материалам диссертации на расширенном научном семинаре кафедры математического и компьютерного моделирования Таджикского национального университета.

Отзыв подготовили доктор физико-математических наук, профессор кафедры математического и компьютерного моделирования Одинаев Раим Назарович и доктор физико-математических наук, профессор кафедры теоретической физика Солихзода Давлат Куват, а также обсужден, одобрен и утвержден на расширенной заседании кафедры математического и компьютерного моделирования механико-математического факультета Таджикского национального университета 13.06.2024 года, протокол № 11.

**Председатель заседания,**

профессор кафедры информационных  
и коммуникационных технологий,

доктор физико-математических, профессор



Нуров И.Дж.

**Эксперты:**

доктор физико-математических наук,  
профессор



Одинаев Р.Н.

доктор физико-математических наук,  
профессор



Солихзода Д.К.

**Секретарь:**

кандидат технических наук, доцент



Одинаев А.Х.

Сведения о ведущей организации:

Таджикский национальный университет 734025, Республика Таджикистан,  
г. Душанбе, проспект Рудаки, 17,

Тел. (+992 -37) 221-62-25; (+992 -37) 227 -15-10, e-mail: [info@tnu.tj](mailto:info@tnu.tj),

Сайт: <http://www.tnu.tj>

Подписи Нурова И.Дж, Одинаева Р.Н., Солихзода Д.К., Одинаева А.Х. заверяю.

Начальник УК и СЧ ТНУ



(ПОДПИСЬ)

Тавкиев Э.Ш.